QUESTIONS
AND
RELATIVE AND COMPLEMENT CLAUSES
IN A BANGLA GRAMMAR

By

Probal Dasgupta

with a Foreword by

Richard Kayne

A JJLL Publication of School of Languages and Linguistics
Jadavpur University
Kolkata
2020

Journal Homepage: http://jjll.jdvu.ac.in/journal/index.php/JJLL

Jadavpur Journal of Languages and Linguistics
ISSN: 2581-494X

Volume 4
Supplementary Issue

Questions and Relative and Complement Clauses in a Bangla Grammar

This dissertation, in the Department of Linguistics, was originally submitted to the Faculty of the Graduate School of Arts and Sciences
by
Probal Dasgupta
in partial fulfillment of the requirements for the degree of Doctor of Philosophy at New York University in the month of February, 1980.
The submission was approved
by
Lewis Levine
(research advisor)

with a Foreword by
Richard Kayne

***

Probal Dasgupta's "Questions And Relative And Complement Clauses In A Bangla Grammar"

Richard S. Kayne
New York University

Probal's 1980 NYU Ph.D dissertation takes us almost as far back in the history of generative syntax as my 1975 *French Syntax*. Both are harder to read now than they would have been forty years ago, as is other syntactic work from earlier decades. But we current syntacticians must be able to go back to that earlier work, to see through the earlier theoretical frameworks in such a way as to allow us to learn from that earlier work, to distil from it that which has essentially remained unchanged.

Probal's dissertation makes use of highly specific transformations of a sort that was common in the first couple of decades of generative syntax. Each transformation is associated with a structural description (SD) and a structural change (SC). Neither SDs nor SCs are in use anymore. Yet in many (though not all) cases, those earlier transformations correspond quite closely to syntactic operations that play a role in current minimalist work.

In reading Probal's 1980 work, we need to transpose it in our minds, to update it to a more current framework. We need to figure out which of the very many ideas in it are correct in essence and what the new form of those ideas should be. In some cases the needed changes may be small, in others larger.

If antisymmetry is correct, in particular in its requirement that all syntactic movement be leftward, then instances of postposing given in the dissertation must be recast (and similarly for instances of postposing found in *French Syntax*). Remnant movement, which goes back especially to work by den Besten and Webelhuth and which is prominent in antisymmetric work, will need to play a more central role in Bangla syntax than anyone could have thought in 1980.

Probal takes into account an impressive number of languages. Now, forty years later, the field of syntax is populated by work on a far greater number of languages. What we have learned from them will further inform our understanding of Bangla syntax, as will, needless to say, important work on Bangla itself done since 1980 by Probal himself and by others. This increase in the number of languages feeding into the field of syntax can justly be thought of as exciting; at the same time it has made the field more challenging. Large numbers of languages are hard to keep track of, hard to keep on top of. Yet evidence bearing on Bangla grammar needn't come only from Bangla, given that all languages are tied to a common language faculty.

Probal's dissertation work bears on all sorts of current theoretical questions. One very general one, which comes up indirectly at many points in the dissertation, concerns the relation between syntax and morphology. To what extent can morphology now be reduced to syntax, a question central to a recent 2020 workshop at NYU (cf. https://wp.nyu.edu/morphology_as_syntax/)?

On the morphological side of things, there are many proposals that Probal correctly makes concerning the decomposition of various complex forms in Bangla. In some cases (switching to English), decomposition is straightforward, e.g. *somewhere* is clearly *some* + *where*, in other cases, less so. For example, is *where* itself bimorphemic, as many would say (*wh+ere*), or is it trimorphemic (*wh+e+re*), with the middle *-e-* found in *there*, but not in *here* (cf. chapter 5 of *Comparisons and Contrasts*)?

Probal has very interesting discussion of Bangla emphasizers, including /o/, as well as very interesting discussion of indefinite-forming /o/. The question arises as to whether or not they are homophones. Both Embick and I (in chapter 7 of *Questions of Syntax*) take the position that the language faculty looks askance at homophones (in particular within the functional vocabulary of a given language). If so, then these two instances of Bangla /o/ must not have identical phonology, which they in fact seem not to, at least to judge by the fact that Probal points out (in the course of generally paying careful attention to Bangla phonology, as he also does to Bangla semantics), to the effect that the emphasizer lengthens a preceding vowel in a way that the indefinite forming one does not. (The question of the validity of 'anti-homophony' also arises with respect to *ki*, and *je*.)

A substantial part of Probal's dissertation has to do with complementizers. Strikingly they can occur in Bangla in non-final position (in a way that to my eye is reminiscent of Basque), with certain restrictions, as usual. Can this property of Bangla complementizers now best be made sense of from an antisymmetric perspective, or not? And was I right to ask, in chapter 10 of *Comparisons and Contrasts*, why English has *that* as a complementizer, but not *this*? If I was, and if the answer I gave is on the right track, then what we call complementizer *that* is really a demonstrative. (Diachronic change(s) affected it without removing its demonstrative character.) If so, could something similar hold of (some) Bangla complementizers?

The question of complementizers is closely related to the question of relative clauses, which are generally taken to look quite different in Bangla, as compared with English. More specifically, it's often thought that English lacks the correlative type (what Probal calls 'left relatives'). Yet my English is fairly happy with *?Whichever man falls in love with whichever woman, they should get married right away*, with *ever*. (Perhaps, then, Bangla 'left relatives' contain a silent counterpart of English *ever*.)

Any analysis of complementizers and relative clauses must now also take into account Rizzi's 1997 paper, as well as, more indirectly, Jayaseelan's of 2001, which leads one to wonder if Bangla, too, might not have IP-internal topic and focus phrases, in Jayaseelan's sense.

As Probal's work so well illustrates, Bangla syntax bears on a wide range of questions of theoretical importance (which go well beyond what I've touched on here). Reading or rereading his dissertation, while keeping an eye on the necessary transpositions to the present, remains well worth the effort.

References:

Besten, H. den & Webelhuth, G. (1987) "Remnant topicalization and the constituent structure of VP in the Germanic SOV languages," abstract, *GLOW Newsletter*, 18, 15 - 16.

Besten, H. den & Webelhuth, G. (1990) "Stranding," in G. Grewendorf & W. Sternefeld (eds.) *Scrambling and Barriers*, Academic Press, Amsterdam, 77-92.

Embick, D. (2003) "Locality, Listedness and Morphological Identity," *Studia Linguistica*, 57, 143-169.

Jayaseelan, K.A. (2001) "IP-internal Topic and Focus Phrases," *Studia Linguistica* 55, 39-75.

Kayne, R.S. (1975) *French Syntax. The Transformational Cycle*, The MIT Press, Cambridge, Mass.

Kayne, R.S. (1994) *The Antisymmetry of Syntax*, The MIT Press, Cambridge, Mass.

Kayne, R.S. (2010) *Comparisons and Contrasts*, Oxford University Press, New York.

Kayne, R.S. (2019) *Questions of Syntax*, Oxford University Press, New York.

Rizzi, L. (1997) “The fine structure of the left periphery,” in L. Haegeman (ed.) *Elements of Grammar*. *Handbook of Generative Syntax*, Kluwer, Dordrecht, 281-337.

Foreword to the 2020 version
Probal Dasgupta

This dissertation, defended in 1979 and submitted to New York University's Graduate School of Arts and Science in 1980 for a Ph.D. in Linguistics, has now been sanitized for serious use. While revising the text, I have been mindful of the need to avoid anachronism. A persistent (but mercifully microscopic) flow of requests for this text over the decades has made me realize the need to make a reasonably error-free version available to linguists in general, not only to the researchers who have individually asked for the dissertation. It is now appearing as a supplementary issue of the *Jadavpur Journal of Languages and Linguistics*.

Earlier users of this dissertation have obtained copies of the original, and tend to refer to its page numbers. In view of the fresh pagination – I take it that users will now regard this new version as definitive – I suggest the following referencing strategy that won't send readers rushing to get access to the original 1980 text. Those of you wishing to pinpoint crucial passages may please mention chapter numbers and example numbers (e.g. "see the discussion between examples (74) and (75) of chapter 6"), which remain unchanged.

What does change, then? I am finalizing this 2020 version in India, after a lifetime of working with linguists who are accustomed to the orthography and punctuation used in Commonwealth countries (apart from colleagues and students in India itself, I'm thinking of friends in Australia, Bangladesh, Indonesia, Malaysia, Pakistan, Srilanka). In keeping with my current practices, I have converted the orthography and punctuation to British norms – retaining the Oxford comma, however. This also means that in a couple of passages where my 1980 original emulated American practices in orthography-adjacent matters (such as "an historical" rather than "a historical") I have switched over to the British norm. Place name conventions have not been anachronistically modernized. In 1979 (and a fortiori at the time when the relevant publications appeared) it was normal to abbreviate Connecticut as Conn. and Massachusetts as Mass.; the references, therefore, retain those forms and do not change them into CT and MA. Since I am publishing in *JJLL*, obviously I'm following their stylesheet on format issues.

I have also tweaked the transcription, minimally. The Ray et al. system used in the 1980 version uses capital N for the velar nasal; this practice collides with the widespread use of that symbol for the retroflex nasal in Hindi-Urdu, Marathi etc. This time, therefore, I have chosen the digraph ng for that role, and made the concomitant decision to show an overt syllable boundary in the very few words where Bangla has an alveolar /n/ followed by a velar plosive /g/, e.g. /kon$gulo/ 'which ones'. When drafting my thesis, I didn't quite notice that I had not made any policy decisions regarding the transcription of Hindi-Urdu examples. This meant that I simply carried over transcriptions by others without comment, including capitalized vowel characters that stand for nasalized vowels – a convention that no reader could possibly have guessed without assistance. Belated apologies; there are no opacities of that kind in the version I am providing this time. All that you need to know to parse the very few Hindi examples is that double vowel letters stand for long vowels.

Turning to matters of grammar and style, while revising the text, I found that some of my indefinite articles instantiated a common Indian hypercorrection: I had repeatedly chosen *a(n)* where *one* would have been appropriate. South Asian languages do not have an indefinite article distinct from the numeral *one*. I have corrected these and any other errors of grammar or infelicities of style that caught my eye, but I was at pains to preserve the flavour of my 1979 writing.

Only two terminological decisions have been revisited, to avert confusion. In my original text I used the term "correlative" to designate determiners and pronouns in the s/t-series, /Se/ 'he or she', /tara/ 'they' etc.; but linguists ordinarily associate the term with the "correlative construction" in which pronouns or determiners of this type work in concert with relative pronouns or determiners; calling a particular class of words "correlative" words turned out to be a stumbling block. I have now switched over to the term "sequent" for that class of pronouns and determiners, and revised the text accordingly. Likewise, readers have found confusing my unusual decision to designate classifiers and measure words as "numeratives" (abbreviated NUMV); I have replaced it with the term "Denominator" (abbreviated DENOM). Readers familiar with my later writings will recognize these choices; the community has yet to converge on standard terms for these notions.

It would have been unfair (to the point of 'rewriting history') if I were to erase or embellish passages that exemplify poor judgment on my part, so I've been careful not to edit them out. For instance, in chapter 5, examples (73) and (74), and what I say about them, exhibit classic symptoms of descriptive confusion. I use the term 'Objective Case marker' without distinguishing the dative from the accusative use of /ke/; I fail to distinguish ingestives from other causatives. These failures, however embarrassed I might be about them, do not stop you today from grasping the facts; nor is there any risk of their misleading you. The few downright mistakes I found – in the reporting of facts or in the description – might mislead, and have therefore been corrected. One place where such correction is non-trivial is the appendix to chapter 3; the extensive revision there, flagged in the note you will find in the appendix itself, affects nothing in the main exposition. I hope there are no other, undetected errors of similar magnitude.

Turning to matters of content, I would like to flag only three significant points. Point one: this dissertation tries to put semiotics on the map as an adjunct to formal linguistics. That I made this attempt, at that early moment of the development of generative grammar, is due to three mentors whose work added the semiotic instrumentation to my toolkit. Ashok Kelkar taught me at Deccan College, Pune. Nicolas Ruwet did not teach at NYU, but gave a guest lecture within weeks of my arrival; during my studies in India I had already read his work. He was a constant presence in the social milieu; I had several one-on-one sessions with him in addition to the important conversations hosted by Michiko Kosaka – a friend who had just finished her coursework by the time I started mine. My third mentor, Ray Dougherty, directly taught me at NYU and was the syntactician on my advisory committee; since I invoke his 1978 work explicitly in the text, I need not say any more in these comments.

The second point to highlight is my taking up, from Michael Helke (who had taught at NYU from 1972 to spring 1975; although I met him only in 1979, his lasting influence on the senior students had been a major factor in the debates raging when I arrived in fall 1975), the use of silent lexemes as a descriptive resource. As readers today are aware, Richard Kayne later made extensive use of this device, which he had independently developed, and which is differently niched in his system. Helke's paper (1979) appeared in a festschrift in Japan and was not widely noticed in North America or Europe (I interrupt these remarks to thank Helke personally, first of all for sending me the typescript of his paper in 1979 and secondly for providing a scanned copy of the printed version in 2019, when I asked him for help). My thesis was probably the first document on record in North America that used silent lexemes explicitly; whether this is of bibliographic importance is for readers to say.

My third and last point is that, on finding that chapter 8 provides only a slipshod and

semi-explicit exposition of my ‘Simplexity’ hypothesis, I have added an appendix, formulating Simplexity with some care, and presenting further examples to give the reader some idea of the kind of follow-up that is still awaited.

Making this version of my thesis error-free, by comparing it with the 1980 original, was a burden shouldered by Malasree Dasgupta. What she had to deal with was two different kinds of fresh typing. Chapter 1 to 5 had been retyped by me, with considerable amounts of ill-considered extensive redrafting on which I had later changed my mind; my request to reverse those revisions meant her having to re-retype those passages. Chapters 6 to 11 had been keyed in by paid assistants relatively unfamiliar with linguistics writing, which meant that she had to deal with “alternation” and “preposing” having been typed as “alteration” and “proposing”. Only four years after her painstaking copy-editing exercise did I at last pick up the thread where she had left off. If this product gives any satisfaction on the typographical front, most of the credit goes to Malasree. I thank Praveen Singh for taking a fresh look at the output of our work and correcting some residual errors. I thank Rajdeep Banerjee for cleaning up the effects of a software glitch that caught me unawares at the last minute.

My 2020 foreword ends here; from this point on, what lies before you is the 1980 text, duly revised. I am grateful to Samir Karmakar for agreeing to publish it as a supplementary issue of *JJLL* and to Richie Kayne for consenting to write some prefatory remarks.

Acknowledgements

If the truth were to be told, these acknowledgements would become longer than the bibliography and like the latter would drown everyone in alphabetical anonymity. So I will arbitrarily limit myself to three sets of direct acknowledgements.

I learnt most of the linguistics used in this research from my fellow students at Sanskrit College, Calcutta, and New York University. Direct contributors to this dissertation include Michael Board, Eileen Fitzpatrick, Joanne Grumet, David Halitsky, Michiko Kosaka, Elaine Marsh, Ngo Thanh Nhan, Lyn Ohira, Dennis Perzanowski, Udaya Narayana Singh, Diane Tone, and Margaret Wade.

Of the professors whose courses I have attended at the University of Pune and New York University, the following have taught me things which bear on this dissertation and have affected its content: Mark Aronoff, Joan Bachenko, Mark Baltin, Martin Braine, John Costello, Ray Dougherty, Ashok Kelkar, Lewis Levine, and Kashyap Mankodi.

My encounters with the following people in the United States and Australia (the country where chapters 4-10 were written) have made a difference to what is in this volume: Keith Allan, Ray Cattell, Ellen Cohen, Robert Fiengo, Jim Gair, Michael Helke, Peter Hook, Norbert Hornstein, Miriam Klaiman, Barbara Lust, Marian Maddern, Geoff Millar, Paul Modini, Stephen Paterson, Peter Paul, Stephen Seegmiller, Ian Smith, and Roger Wales.

Many of the people mentioned have said things which I disagree with; others I think I agree with but may have understood incorrectly; most of the people have simply made me aware of facts and ideas which were new to me. None of them is responsible for my failure to grasp what they have taught me. I feel thankful to them all and to the many unnamed people whose impact on this research has been indirect; especially to my parents, to whom I hereby dedicate this volume.

Table of Contents

Chapter 1

INTRODUCTION

1.0 The how and what of this work

Long before it became fashionable in the social sciences to adhere to non-rationalistic philosophies of science, Whitehead (1933: 91) declared that our reasonings 'grasp at straws for premises and float on gossamers for deductions', in the process of affirming that the emotional and customary underpinnings of rational inquiry are primary and that its self-conscious technical structure is always less well understood by the users of that structure than they realize. In my view, nothing has taken place or been said in the intervening half-century to weaken Whitehead's point. So I will introduce my work without hiding its gossamer.

This study of the grammar of the Indic language Bangla deals with four clause classes which I will call Relative Clauses or R, Constituent Questions or Q, Complement Clauses or C, and Yes-no Questions or Y. Assuming certain definitions (see the end of this section), the present work offers a method of generating Bangla R, Q, C, and Y. The proposed method generates these clauses in a way that accounts for many morpho-syntactic and semantic (dis-)similarities among the four types of clauses. For some answers to *ad quem* questions at this point see section 1.1 which is entitled 'the why and wherefore of this work'.

Tools (assumptions, rules) accepted in the literature of lexicalist grammar are the gossamer of this study. Such acceptance is justified by empirical studies of English, mostly, and of other languages as well. When establishing or reinforcing the rules and assumptions, those studies relied, in turn, on other gossamer. Some examples follow. This study makes crucial use of the concepts Complementizer, Determiner, Specifier, +OPEN, the rule-governed binding of +OPEN by an antecedent, and the assumption that a construal rule central to the semantic analysis of R and Q relates the Complementizer of an R or a Q to some phrase node inside the clause. I have taken and in part adapted these ideas from studies of English, or from theoretical papers based mostly on English data, which present some results of a decade of attempts by many linguists to strike a balance between the use of transformational and of lexical descriptive devices. Let me clarify this point with reference to each of the ideas mentioned.

The concept of COMP (Complementizer) as part of the rule S′ → COMP S is due to Bresnan 1970, a paper which weighed the merits of a transformational treatment and a lexical treatment and argued for lexical insertion of complementizer words into trees. Bresnan's rule given above, using bar notation the way it does, tacitly claims that COMP bears the SPEC (Specifier) relation to S (Sentence).

Let us turn to SPEC. This concept first appeared in Chomsky 1970 where the relation of AUX to V′ and that of DET to N′ were viewed as special cases of the relation of a SPEC to its sister node in an X′ expansion. Chomsky 1970, written in 1967, was the first discussion of the impact of the lexicon, introduced in Chomsky 1965, on generative linguistic theory. The discussion stressed just one implication of having a lexicon: that transformational analyses would now need defence against lexicalist alternatives and that such defence would be difficult or impossible in those cases (e.g. English nominalizations) for which the lexicalist alternative could combine X′ notation's cross-categorial generalizations with the use of phrase structure

rules and lexical features to express particularities. But there is more to lexicalism[1] than the ability to do what transformational analysis does and to do it better. The apparatus of bar notation suggested making generalizations which the transformational methods of the early sixties would have not led one to think of. In particular, the SPEC concept as it stands today embodies two such generalizations.

The first of these generalizations is that expressed by Jackendoff's (1977) notion of DET which includes both the NP (Noun Phrase) introducers like *the* in *the toy* that were called Determiners before and the AP (Adjective Phrase) introducers like *this* in *this big* that have also come to be called Determiners. (Jackendoff has DET as a super-category; he calls DET in NP "ART" and DET in AP "DEG".) This new generalization, a by-product of the X′ machinery Chomsky built for the nominal and verbal systems, corresponds to nothing transformational. There never was nor would be any question of inserting either kind of DET transformationally. It is in the broader sense of DET, for both NP and AP, that I use *DET* in this study.

The second generalization encompasses the broadened DET as well as COMP and AUX in the notion of SPEC. See Halitsky 1975 for one paper which makes explicit Bresnan's (1970) proposal that COMP be considered a case of SPEC.

Moving on to the assumption that a rule central to the semantics of R and Q at least in English relates the COMP of the clause to some clause-internal phrase node, this assumption has been common currency since Baker 1970 and Bresnan 1970. The word *some* is to be understood here as 'at least one'. Baker 1970 had particularly interesting things to say about cases involving more than one clause-internal phrase node; e.g. in *Who bought which book?* the phrases *who* and *which book* must both be related to COMP (to Baker's "Q") without invoking syntactic preposing of *which book*. The matching mechanism proposed by Baker for cases like *which book* in English must operate, I show in this volume, throughout all R and Q clauses in Bangla. Bangla does not prepose its relative or interrogative phrases.

Consider, finally, the assumption of rule-governed binding of +OPEN by an antecedent and the concept +OPEN which figures in this assumption.

The English WH morpheme occurs in a set of words (*which, who, where,* etc.) which do both relative and interrogative work. Kuroda 1968 is still the best account of how English WH unites these two functions. Other languages like French, Russian, or Bahasa Indonesia also manifest this syncretism in varying degrees. One may set up a construct OPEN in universal grammar and postulate that English WH, French QU, etc. are realizations of its plus value. One element in the definition of OPEN must be that a +OPEN position invites binding by an antecedent. A traditional case is, say, the antecedent *those* binding the relative *who* in *those who arrived*. The present study claims that binding also occurs in questions. Thus, in *Who arrived?,* the WH position is bound by a semantic zero which, being zero, leaves the kind of gap that leads the clause to be a question – to invite an answer. This image of a gap that is either closed by non-zero binding or left open by zero binding underlies the choice of the term OPEN. Besides, logicians use the term *open sentence* for a notion allied to my *open clause.* A logical 'translation' of an open clause in my sense would, in fact, be an open sentence in the

[1]Lexicalism is adherence to Chomsky's (1970) 'lexicalist hypothesis' that, in general, some grammatical generalizations are better expressed by cross-categorial or other bar-notational rules acting in concert with lexical specification than by exception-ridden transformations. Chomsky 1970 discussed English nominalization as an example.

logicians' sense.

I have clarified most of the concepts which figure in the principal claim of this volume, the Open Clause Hypothesis. Here is a preliminary statement of the claim.

One primary factor, a +OPEN feature in the COMP node, plus two secondary factors, the value (+ or –) of the INT (Interrogative) feature in COMP and the presence or absence of +OPEN Determiners in the S sister of the COMP, jointly decide that an S′ will be an Open Clause and whether that S′ will be R, Q, C, or Y. In terms of my grammar these clause types have the following properties which distinguish them from other clause types.

An R clause, or an R, contains one or more +OPEN –INT Determiners.[2] A Q contains one or more +OPEN +INT Determiners. A C contains a +OPEN –INT Complementizer. A Y contains a +OPEN +INT Complementizer. If we think of the items listed as being 'flags' marking these clause types, it may be said that each clause type has its own 'flag' and lacks those of the other three. Thus, a Q must not contain a +OPEN –INT Determiner or any sort of Complementizer. But all four clause types have a COMP node marked +OPEN. C and Y must, and R and Q must not, have a word in COMP position. How the nodes, words etc. get there and what work they do on various linguistic levels is too large a question to exhaust here. The answers to it will vary from one language to the next within the range of application of the Open Clause Hypothesis – an as yet undetermined range. One language which may fall outside this range is Japanese, which has no words that function as relative words. Both a language like English, in which a WH morpheme manifests +OPEN (both +INT and –INT) in DET position, and a language like Hindi, which has a J morpheme for +OPEN –INT and a K morpheme for +OPEN +INT, fall within the range. See chapter 9 for some indication of the limits of the hypothesis.

Chapters 7-9 present an analysis of Bangla open clauses. Chapters 2-6, which prepare for this analysis, concern themselves with local problems of Bangla structure, seldom studied by syntacticians and to some extent amenable to solution even without the findings about higher clause structure which are given in chapters 7-9.

I will now give a brief guided tour through the volume and then try to make precise what I mean by R, Q, C, and Y in Bangla and what relations I establish among them.

Chapters 2-4 build a skeletal, overall Bangla grammar. Chapter 2 does the morphology of N and V, including some phonology, without which the morphology of Bangla verbs is nearly unintelligible. Chapter 3 establishes some phrase structure expansions containing N and V. The analysis of pronouns in chapter 3 shows that some pronouns are N while others are DET Z sequences, Z being some noun with no phonological realization. This point is very important. It permits us to localize the Open (relative /interrogative) element of relative and interrogative phrases entirely in the DET node, even when we have, not a phrase like *kon ghOr* 'which room?', but a pronoun like *ke* 'who?' – the latter is an NP with the DET *ke* preceding a silent N. Chapter 4 pursues the verb into the infinitive system and distinguishes the infinitive form, V, from the participle-gerund form, N as a gerund and A as a participle.

Chapter 5 further studies infinitives, now in relation to the problem of open clauses. Bangla does not permit relative infinitivals (English does: *a knife with WHICH to cut diamonds*) or interrogative infinitivals (like English *WHAT to do*). Bangla infinitivals are also

[2]Containing Determiners or Complementizers means containing DET and COMP nodes that dominate actual words rather than just null elements.

demonstrably not clauses. Chapter 5 links these facts via the idea that for something to be able to be a relative or an interrogative construction that thing has to have the structure [S COMP], which an infinitival, being a VP, does not have. This idea presupposes the existence of complementizers and their importance in the grammar of R and Q clauses – two arguable points.

Chapter 6 demonstrates the existence of Bangla complementizers and examines their syntagmatic and paradigmatic ties with two other categories – emphasizers and conjunctions.

That complementizers have something to do with R and Q clauses is shown in chapters 7 and 8. Chapter 7 begins by tabulating deictic and pronominal forms and establishing a morphological contrast between + and –OPEN and, within +OPEN, between + and –INT. The determiners whose presence in an R makes it an R are +OPEN –INT and those whose presence in a Q makes it a Q are +OPEN +INT. These feature bundles show up as J and K respectively. Now, not only DET can host the morphemes J and K. COMP can, too. When it does, the resulting clause is a C containing a complementizer *je* or a Y containing a complementizer *ki*. Therefore, the COMP node can carry the feature bundles characteristic of the open DET nodes of R and Q. Now, suppose COMP is present as a node carrying those features, albeit not dominating any morpheme, in every R and Q. Chapter 7 shows that this empty COMP analysis of R and Q does better than obvious alternatives. Chapter 8 offers a full study of structures containing indefinite words (which are interrogative words plus a constant increment), relative words, and interrogative words.

Chapter 9 extends the analysis to C and Y and states the principal claim of this inquiry, the Open Clause Hypothesis. Chapter 9 also lists some problems that arise in or just beyond this study but are not solved here.

Both the general approach to linguistics and many details of analysis in this study are more akin to works by such authors as Bresnan and Brame than to works by authors like Chomsky and Lasnik. In many cases the former and the latter differ only in attitude or terminology. But where I have found empirical differences, I have generally found the former to be closer to the truth. Some of these cases are explored in chapter 10, which does not limit itself to Bangla data.

Throughout the study, but especially in chapter 10, small or medium-sized analyses follow one another without adding up to a gapless theory of Bangla or of language. This diffuseness of inquiry seems unavoidable in our present state of ignorance. In order to compensate a little for this diffuseness, I will now define my key terms. Chapter 11 will list the final forms of rules postulated. Thus, the study begins and ends with efforts to systematize.

A relative clause or an R in this Bangla grammar will, throughout, be a sentence which contains one or more relative phrases and occurs to the left of its antecedent/s, where a relative phrase is one that contains a word beginning with the morpheme J. The first four words in the following sentence are a relative clause: *tumi je juto kinecho Se juto manuSer paYe hOY na* "you which shoe/s have-bought that shoe/s human-GEN foot-LOC is not" 'The shoes which you have bought do not fit human feet'.[3] In this example the antecedent *Se juto* 'those shoes',

[3]Double quotes indicate word for word glosses and single quotes translations. The transcription for Bangla, following Ray et al. (1966), uses capital letters as follows: E O low, S palato-alveolar, N velar, T D R retroflex, Y W mid, M nasalization of preceding nucleus; we depart slightly from Ray et al. in that we use the digraph *ng* for a velar nasal (instead of their

as required, follows the relative clause. Bangla also has 'right relative clauses', discussed in chapter 8, where the antecedent does not follow, but precedes. The present study does not try to analyze Bangla right relative clauses and therefore excludes them from the definition of R. See chapter 8 for reasons for this exclusion.

I here define a constituent question or Q as a sentence containing one or more interrogative phrases, where an interrogative phrase is one containing a member of the K-Word series *ke ki kEno kOkhon...* 'who? what? why? when?...' As chapter 7 says in detail, I exclude exclamatory constituent questions and, without discussion, echo questions. An example of a Q: *tumi kon dokane giyechile?* "you which? shop-LOC went" 'Which shop did you go to?' Someone who had not heard this question might say: *%ke% kon dokane giyechilo?* 'Which shop did %who% go to?' – using *%%* to indicate the echo question word. My grammar cannot account for *%ke% kon dokane giyechilo?* with its peculiar intonation and violation of what chapter 8 calls the simplexity principle. To the best of my knowledge, no student of any language has satisfactorily dealt with echo questions yet.

A complement clause or a C, for present purposes, is a sentence which contains *je* functioning as a complementizer. The first four words of the following sentence are a C: *dokan je bOndho thakbe kawke Se kOtha bolo na* "shop that closed will-be anyone-OBJ that thing tell not" 'Don't tell anyone (the fact) that the shop will be closed'. This definition excludes *je* + S constructions like the last four words of *kawke bolo na je dokan bOndho thakbe* "anyone-OBJ tell not that shop closed will-be" 'Don't tell anyone that the shop will be closed'. The reason is that *je* in such occurrences is a Conjunction and not a Complementizer; see chapters 6 and 9.

A yes-no question or a Y, in this volume, is a question which by virtue of its very format requires a yes or no answer. Thus, *ram ki jabe?* "Ram whether will-go" 'Will Ram go?' is a Y. In contrast, the question *ram jabe, na, jabe na?* "Ram will-go, or, will-go not" 'Will Ram go or not?' permits a (slightly infelicitous) yes or no answer because of the choice of words; but without changing the format we could alter the question to *ram jabe, na, thakbe?* "Ram will-go, or, will-stay" 'Will Ram go or stay?' which precludes a yes or no answer. Thus *ram ki jabe?* is, unlike *ram jabe, na, jabe na?*, a Y.

The relations which this volume establishes among the four clause types may be summed up as follows:

	in Comp position	in one or more Det positions
+OPEN, –INT word	C	R
+OPEN, +INT word	Y	Q

As the above paradigm indicates, the following groups are natural classes: C R Y Q; C R; Y Q; C Y; R Q. The warrant for such a classification can only be similarities and differences of distribution, function, and internal structure. Some examples of the relevant similarity and

---

capital N). For Hindi-Urdu, we use the same characters for the same purposes, plus N for a retroflex nasal, double letters for long vowels (ii, aa etc.), and the morphophonemic glyph @ for aa (masculine singular) ~ ii (feminine) ~ e (masculine singular oblique or masculine plural).

difference data follow.

R and Q form a natural class: an R may have several relative phrases and a Q several question phrases: *ke kon dokane giyechilo?* "who? which? shop-LOC went" 'Who went to which shop?'; *je jOto pORaSuno koreche Se tOto boka* "who how-much study has-done s/he that-much stupid" 'The more one has studied, the dumber one is'. In contrast, a C cannot contain more than one *je* 'that ', nor a Y more than one *ki* 'whether': *dokan ki khola chilo?* "shop whether open was" 'Was the shop open?'; *dokan je khola chilo ami Se kOtha jantam na* "shop that open was I that fact knew not" 'I didn't know the fact that the shop was open'. Thus, C and Y form a natural class distinct from the R-Q natural class.

R and C form a natural class. In the grammar of both, J-Words play a major role, J being the morpheme for [+OPEN, –INT]. A clause like *ram je kOtha bolchilo* is in fact ambiguous between an R reading '(the thing) which Ram was saying' and a C reading 'that Ram was speaking'. The R reading is appropriate in *ram je kOtha bolchilo Se kOthaTa Ekdom notun* "Ram which thing was-saying that thing-CLA quite new" 'The thing which Ram was saying is quite new', where the classifier *Ta*, with its definiteness effect, makes it difficult to think of the C reading. The sentence *ram je kOtha bolchilo SEm joduke Se khObor dEY ni* "Ram that word was-saying Shyam Jadu-OBJ that information has-given not" 'Shyam has not told Jadu that Ram was speaking', in contrast, makes the C reading much more natural. In both cases the word *je* links up with the word *Se*. In the sentence where the R reading is more natural, *Se kOthaTa* 'that thing' binds the relative phrase *je kOtha* 'which thing'; in the other sentence, *Se khObor* 'that information' binds instead the complement clause *ram je kOtha bolchilo* 'that Ram was speaking' as a whole.

Q and Y form a contrasting natural class. They both highlight K-Words, K being the realization of [+OPEN, +INT]. The string *ram ki kOtha bolchilo?* is ambiguous between "Ram what thing was-saying" 'What was Ram saying?' and "Ram whether word was-saying" 'Was Ram speaking?' – a Q and a Y reading, respectively. The word *ki*, when it means 'what', focuses the question on the phrase *ki kOtha*, and, when it means 'whether', spreads the question over the entire clause.

I have no raw data showing R-Q-C-Y to be a natural class. This larger natural class is obtained by inference from the smaller ones.

1.1 The why and wherefore of this work

Why is it important to establish and study the Bangla open clause paradigm whose rows and columns form the natural classes just mentioned? Some broad and some narrow answers follow, in that order.

This study claims that the COMP word of a C and the relative DET word/s of an R share a grammatical feature bundle, +OPEN –INT. This claim, if correct, offers a general or principled account of the similarity between the C complementizer and the R pronoun-determiner system in many languages: Sanskrit (COMP and DET *yat*), German (COMP *dass*, DET *das*), French (COMP and pronoun *que*), Russian (COMP and pronoun *shto*), etc. This similarity, though a widespread phenomenon, has not been widely noted as a phenomenon that students of universal grammar need to explain in terms of linguistic theory. The present study notes the problem as it appears in Bangla and proposes a language-particular solution based on universal constructs like +OPEN and COMP, thus building some of the general principles on

which, some day, a solution to the larger problem may rest.

Within my experience Bangla is a unique language in the extent to which its R, Q, C, and Y in their very morphology reveal the similarities and differences among them. Thus, Bangla is an optimal arena for the study of the open clause system.

Another claim made in this study is that the +OPEN DET of an R and the +OPEN DET of a Q share the + value for OPEN but differ in their values for INT. This study specifies the difference as follows. The +OPEN –INT determiner in a relative phrase is bound by an antecedent. The +OPEN +INT determiner of an interrogative phrase is bound by a special semantic zero which, if you wish, is a 'Gap left open!' announcement. This claim, if correct, offers a principled account of how questions are questions, of how they invite answers, and at the same time leaves the possibility open that some languages, like English and French, might have a single morphemic realization (English WH, French QU) for +OPEN, while other languages, like Hindi and Bangla, may pick one morpheme (Hindi-Bangla J) for +OPEN –INT and another (Hindi and Bangla K) for +OPEN +INT.

As these remarks indicate, the Open Clause Hypothesis of this study is intended for two levels: general linguistics and Bangla grammar. But the fact that this study stresses the morphological side of the problem makes it necessary to ask how the hypothesis is to be applied to languages with morphological systems unlike the system in Bangla. For example, Japanese constructions which render relative clauses of other languages are clause-like structures with the 'relative phrase' sites left blank. There is no set of Japanese relative words. Thus, it is not clear that the Open Clause Hypothesis applies to this language which lacks +OPEN –INT determiners. My inclination is to exclude it from the scope of the hypothesis.

It is less obvious that German should be excluded. German has a set of words, D-Words, which combine a relative and a definite function: *der, das, die,* etc. The Open Clause Hypothesis does not predict such a situation. But it does not predict the contrary either:

Open Clause Hypothesis, General Version

Given +OPEN 'to be bound', +INT 'by semantic zero', and –INT 'by an antecedent', every S′ whose COMP is specified +OPEN (every open clause) must take one of four mutually exclusive options:

R: have a +OPEN –INT DET in a phrase in S and construe this DET with COMP

Q: have a +OPEN +INT DET in a phrase in S and construe this DET with COMP

C: have a +OPEN –INT COMP word

Y: have a +OPEN +INT COMP word

The Open Clause Hypothesis does not contradict the requirement of German morphology that +OPEN –INT determiners be, in addition, +DEF (definite). In fact, a study which does for German what this study does for Bangla may unearth reasons for saying that +OPEN –INT is universally +DEF. That would enrich, not kill, the Open Clause Hypothesis. Just as English WH is always +OPEN but, when relative, also –INT and, when interrogative, also +INT, so also the German D is entitled to be always +DEF –INT and, when true-definite, also –OPEN,

and, when relative, also +OPEN. To put it in prose, just as the English open element WH combines an interrogative with a non-interrogative function, so also the German definite non-interrogative D element may combine an open with a non-open function.

I turn now to a narrower reason why this study is important. There are many cases where existentially or universally quantified expressions derive from an open word. Consider English *somewhere* and *wherever* from *where*, German *irgendwo* 'anywhere' from *wo* 'where', Japanese *dare ka* 'someone' from *dare* 'who', etc. The semantic composition of such words has yet to be investigated properly. The present volume makes a start by relating Bangla indefinite words like *kon-o* 'some, any' and *kOkhon-o* 'some time, ever' to interrogative words like *kon* 'which' and *kOkhon* 'when' both in formal grammar and in notional grammar. Such an analysis becomes possible because the larger description of open clauses, including Q, provides some initial idea of the semantics of interrogative words. The tentative conclusion is that quantification is a composite thing and that one of the components of a quantified expression is a +OPEN element, logically a variable. This makes the theory of open clauses more fundamental than the theory of quantification and suggests a reversal of priorities in linguistic semantics. Writers like Chomsky and Lasnik have assumed that the interrogative WH in English is logically a quantifier. The present study shows that +OPEN (which is WH in English) is, on the contrary, a variable, and forms part of existentially quantified expressions like Bangla *kotha-o* "where-some" 'somewhere'.

Finally, a point which is neither broad nor narrow: without a study like this the syntax of Bangla would remain mostly undescribed. As Chattopadhyay 1976a, the only generative thesis on Bangla which I have had access to, observes, there is no overall analysis of Bangla syntax in any framework, traditional or modern. The gap needs to be filled.

1.2 The choice of lexicalist grammar

Some readers may wish to know why this analysis uses lexicalist methods. At the cultural level, one may ask why an Indian language is not being analyzed within the classical Indian grammatical tradition associated with such authors as Panini. On the level of ideologies within modern linguistics, a reader who is aware that most syntactic work on modern Indian languages has taken place within the research tradition of Generative Semantics (GS) may wonder why I have rejected GS without discussion.

The cultural question cannot be tackled without delving into the history and sociology of linguistic and other scientific inquiry in such third world countries as India. The Paninian approach has not been applied to modern languages, especially at the syntactic level. One might argue on political grounds that such methods should be used; I would not oppose that view. Not having used them, I cannot tell whether Paninian grammatical and allied logical methods offer appropriate concepts for the study of the Bangla open clause system. Maybe they do, and maybe they are better suited to the purpose than the methods used here. Current methodological study of Paninian grammar (Misra 1966) will lead, one hopes, to a revival of the best of the traditional methods. But there is no community of students of modern languages who now use Paninian methods. So, I have had less access to these methods than to the grammatical research methods of the Anglo-American world.

Grammarians of Bangla have traditionally favoured western over Sanskrit-based methods. Ever since the study of Bangla grammar began in the eighteenth century, Bangla

grammarians, like other Bengali scholars, have taken their tools of analysis from Sanskrit-based and English-based research traditions, with increasing emphasis on the latter. Present-day school grammars confine their reliance on Sanskrit-based notions to the morphophonology of Sanskrit loans in Bangla, the rest of grammar being set in a framework adapted from English traditional grammar. The progress of Anglicization in this sphere has gone unchecked by any revivalistic emotions which particular Bengali grammarians may have felt. The straws at which we have grasped for premises, to revert to Whitehead's image, have been western.

Maybe one can take comfort in the fact that the 'tradition' of English traditional grammar is itself Greco-Latin and the fact that the consensual project of modern linguistics has been to avoid all merely traditional models too closely geared to the structures of particular classical languages and to build a method of analysis which is equally applicable to all known languages past and present. The metropolitan object of study of modern linguistics since about 1960 has been English, for well-known social and political reasons. As a result, enthusiasm to free linguistic methods from an overly classical bias has ended up occasionally producing models that work so well for English that they work less well elsewhere. However, modern linguists do wish to create a universal set of methods and concepts. So there has been an increasingly successful effort to resist the new imperialism of English in the world of linguistics. Success in this effort is perhaps the best hope for students of modern Indian languages who want truly bias-free principles of linguistic analysis.

In my view, resistance to the imperialism of English will be most effective if instead of sterilely rejecting concepts and assumptions as 'contaminated' by English or Latin or Arabic we try to overcome old bias by adding many new biases from less well studied languages and compelling all biases to work together in a general theory. The present study is, among other things, such an exercise in setting up a 'coalition' of biases.

But modern western linguistics itself is not a unitary tradition. It includes disagreeing groups of scholars. The particular tradition exemplified here is known as lexicalist since it assumes the lexicalist hypothesis of Chomsky 1970 discussed in section 1.0 above. In contrast, Chattopadhyay 1976a is a study of C and R in Bangla which uses the approach known as Generative Semantics (GS), an approach widely used in research on Indian languages. It makes sense to ask why I did not choose GS when I undertook my research. Other things being equal, the desirability of smooth communication with other workers in Indian linguistics would have dictated the choice of GS.

Other things, however, were not equal. I needed an approach which would make it easy to state, for example, that J and K occurred in two SPEC positions – the COMP position in S′ and the DET position in NP or AP. But GS does not have or allow the notion of SPEC. A basic GS assumption is that COMP, unlike DET, is not base-generated but transformationally inserted. Partly for this reason, Chattopadhyay's (1976a, b) work, which strikes me as a good application of GS to her data, fails to relate R to C although she analyzes both.

Another reason why I did not select GS was social. While I was preparing for the study, basic research in GS stopped and the intellectual leaders of the movement left for other styles of research which held no promise for the analysis of open clauses in the near future. It was clear that the lack of depth in GS theories about open clauses would not be remedied by work in progress and that, if I adopted GS, I would not benefit from exposure to the sort of growth of theory through constant critical exchange between scholars in the field which is the life of a scientific community.

There were corresponding positive reasons why I chose lexicalism. The theoretical core

of lexicalism, i.e. the X-bar (X′) theory of syntactic categories, offered a way to express the generalizations about COMP and DET which interested me. As explained in section 1.0, lexicalism has a concept of SPEC which presupposes that COMP is to S′ what DET is to NP and AP and thus predicts that there will be similarities between the role of COMP in S′ and the role of DET in its mother phrase. These similarities, it so happens, include the similarities between C and Y on the one hand and R and Q on the other hand which the present study focuses on. Thus, lexicalism is extremely well suited to my problem.

I turn now to a social reason for selecting lexicalism. Lexicalism underwent rapid development during the period of research reported here. This development, which contemporaries see in terms of the outstanding disagreements between Chomsky and Bresnan, but which can also be viewed as the forging of partly reconcilable theoretical tools by a variety of lexicalists, fortunately for me focused on the grammar of open clauses, both in general and in English. Thus, lexicalism in its concrete embodiment, as a set of controversial and, therefore, sharply formulated analyses with special reference to open clause phenomena, met the needs of my research project. If critical readers deem the present study successful, such success supports lexicalism as a valid approach to linguistics, without, of course, showing it to be THE valid approach.

1.3 Strict Lexicalism (SL) and Mixed Lexicalism (ML)

All and only lexicalists accept X′ theory and agree that distinguishing morphological and syntactic from semantic factors is worthwhile. But lexicalism is disunited. Theorists have tended towards the two poles named in the section heading: SL and ML (these are designations I have chosen for these communities, not self-descriptions). SL scholars favour taking maximum advantage of lexical descriptive devices and phasing out reliance on syntactic transformations and especially on movement rules. ML scholars prefer to keep using movement rules and show an increasing penchant for tailoring linguistic descriptions so as to maximize the ease with which one can translate from syntactic structure into Fregean predicate logic, a penchant which was once the hallmark of generative semanticists. The two sets of scholars cluster around Chomsky and Bresnan who have emerged as charismatic leaders of ML and SL respectively. Their emergence as such and their intellectual rigour seem to me to add intensity and give direction to the energies of investigators. If Whitehead was right about the primacy of emotional underpinnings in science, then the role of charismatic leadership in academia is not a wholly negative one. For example, I was in the fortunate position of personally wanting to work on what the heart of the SL-ML debate made it timely to examine. So, I welcomed and was not hampered by the direction in which the charisma of Chomsky and Bresnan propelled mainstream work. It was not hard for me to take sides. Where my findings bore on the SL-ML debate, they supported SL. Chapter 10 reports a few forays made into the main issues of the debate, going out of my way to find out if the debate could be decided on grounds independent of the bulk of my investigation. The studies in chapter 10 also support SL and disconfirm ML.

However, my main concern has been truth, not ideological victory for some side. At several points in my inquiry I have postulated movement rules. I suspect that they may need to be restated as lexical interpretive rules. But if it turns out that they must indeed stand uncorrected, I will not regret that I have been proved right and pure SL wrong. I am sure, too, that many readers will wish to use this volume without bothering about the SL-ML debate,

especially readers whose own work happens to differ in focus from the interests of the charismatic leaders (of whom linguistics has many) and who therefore feel that such leadership does not encourage but obstructs scholarly work. I have tried to keep such readers in mind while writing, though with little success. The concepts of modern linguistics have grown up in the midst of debates and seem to need more debate for further development. The chapters which follow are full of polemic. I hope only that the arguments are formulated cogently and civilly, and thus participate in the general rise in standards of civility and intellectual care in the last decade or so of generative linguistics.

1.4 Summary

Let me recapitulate the principal hypothesis as it applies to Bangla. A +OPEN (relative or interrogative) phrase in a Bangla constituent question (Q) or relative clause (R) is not moved into Complementizer position, but nevertheless receives its interpretation via the COMP node, which must be +OPEN and empty. In the other two types of open clauses (of clauses with a +OPEN COMP node), yes-no questions (Y) and complement clauses (C), the only +OPEN node in the clause must be a non-empty COMP.

Semantically, an Open position in a clause is a variable bound by (i) an actual antecedent outside the clause or (ii) a semantic zero. Case (ii) applies to +OPEN +INT elements, those in Q and Y. Bangla spells out +OPEN +INT as the morpheme K. Case (i) applies to +OPEN –INT elements, those in R and C. The Bangla +OPEN –INT morpheme is J.

Bangla grammar yields distributional and structural dis/similarities among the four types of open clauses by means of rules that introduce and process J and K in COMP and DET positions. Thus, all open clauses have a +OPEN Comp node, and different open clauses contain +OPEN words in different positions:

	in COMP	in DET
+OPEN +INT i.e. K-Words:	Y	Q
+OPEN -INT i.e. J-Words:	C	R

The simplicity of this schema rests on two facts. First, even in an NP like *ke?* ‘who?’, the +OPEN element inheres in the DET node. Second, the Bangla infinitival, which is a VP and thus has no COMP or DET, does not participate in the system: a Bangla infinitival cannot be questioned or relativized. It will be convenient to work on facts like these first so that the open clause system can then be described quite simply.

Chapter 2

OVERALL BANGLA GRAMMAR: PHONOLOGY AND SOME MORPHOLOGY

2.0 Strategy and tactics

Chapter 1 posed a descriptive problem in Bangla grammar, the problem of relating R, C, Q, and Y, the four types of open clauses, to each other. To tackle this problem would have been easier than it in fact is if an adequate outline grammar of Bangla had already been written, in whatever framework, providing at least basic notions of structure for the major constructions. But no such outline exists, as Jayanti Chattopadhyay (1976a, b) points out in the only studies of Bangla syntax or semantics I have had access to which touch on any of the constructions studied here. Chattopadhyay managed without an explicit outline grammar within which to frame her proposals about C and R. She seems to have assumed the validity for Bangla, *mutatis mutandis*, of Generative Semantic analyses of such Indic languages as Hindi. As mentioned in chapter 1, I do not share her reliance on GS. So, I will stick my neck out and proceed to offer an overall analysis of Bangla within which one may assess my proposals about open clauses.

Subject-Object-Verb is the preferred word order in Bangla. Hinds (1974) and Schwartz (1972) argue that in such ('SOV') languages the S (Sentence) node directly dominates the argument NPs and the V, unmediated by any VP. Their arguments, which are lengthy and will not be repeated here, work for Bangla. So, I will assume without discussion that in a Bangla finite clause the V does not form a VP with the phrases it governs.

The sort of Bangla described here is a standard variety spoken in urban West Bengal in general and Calcutta in particular. I hope that, but do not know if, some of this material will help describers of related dialects and languages. I also hope that this work will help design Bangla courses and textbooks.

Chapters 2-4 sketch some Bangla structure in preparation for the main discussion in chapters 7-9. Chapters 5-6 link the sketch with the discussion.

At several points, principled choice between alternatives will prove impossible. I will face this problem by making explicitly arbitrary choices, hoping that they vitiate nothing important.

The present chapter studies basic phonology and two aspects of morphology: declension and conjugation. Since AP does not decline, the discussion of declension deals with NP – nouns and pronouns are treated together. As mentioned in section 1.3, I need to show that those pronouns which have to do with open clauses (*je* 'who', *ke* 'who?', etc.) localize their grammatical features (OPEN, INT, etc.) in the DET node; this point is made in chapters 2 and 3. The present chapter stresses that pronouns, unlike bona fide Nouns, can be reduced to bundles of grammatical features. Chapter 3 argues that some pronouns, like *je* 'who', are DET Z sequences where all the overt morphic material (*je* in this case) belongs to DET, Z being an unpronounced Noun.

At the time of writing, Pabitra Sarkar's 1975 Chicago University dissertation was not yet available for consultation. I have relied on Chatterji 1968 and Ferguson 1945 as amplified by Ray et al. 1966 and my own work (1982 [a revised version of writing done in 1972], 2001 [a revised version of writing done in 1977; in the bibliography, this item is – following the original text of this dissertation – flagged as 'forthcoming in 1981', with clarification and exact

references]). Most of what is said here about phonology and morphology is uncontroversial. The little that is not has no relation to the claims made in this dissertation about clause structure.

2.1 Rudimentary phonology

Like Chomsky and Halle 1968 I assume that phonological rules map abstract lexical representations on to concrete phonetic representations but that the latter are not yet as 'concrete' as a physical signal – thus, I am using Chomsky and Halle's distinction between the fully concrete 'physical phonetic' level and the close-to-concrete 'systematic phonetic' level. Following Chomsky and Halle, I will use structuralist phonemic transcriptions to approximate to the systematic phonetic level, ignoring such details as the nasalization of vowels after nasal consonants. A brief sketch follows of the sound system of Bangla at a level of abstraction which is phonemic for a structuralist like P.S. Ray (Ray et al. 1966) and systematic phonetic for Chomsky and Halle.

The language has seven vowels, *i e E a O o u.* Of these, *O o u* are back rounded, *a* back unrounded, *i e E* non-back unrounded, *i u* high, *e o* mid, *E a O* low *[a 2020 addendum: the prevalent feature system today bifurcates these vowels, informally termed 'low' into the 'true low' vowel /a/ and the 'non-low, non-high, non-ATR' vowels /E O/; none of this was of course on the shelf in 1980; readers will, no doubt, read my terminology in its proper context]*. The nasalized vowels *iM eM EM aM OM oM uM* and the alpha-back alpha-round low vowels *E O* occur only in the initial syllable of a morph, except in unassimilated loans like *restoraM* 'restaurant' or *kOmpETibl* 'compatible'. (There are other exceptions, like *SukkhoM* 'delicate'.)

Corresponding to *i u e o* there are glides *y w Y W,* which in one analysis contrast with the former in post-vocalic position: *day* 'midwife', *dai* 'responsible', *dewle* 'bankrupt', *deule* 'in a temple', *daY* 'responsibility', *dae* 'due to responsibility', *Saoner paWna* 'money owed to Shaon'. An alternative analysis, which interposes glides in words like *dayi* 'responsible', *dewule* 'bankrupt', *daYe* 'due to responsibility', can be reformulated as an account that minimizes semivowel occurrences by claiming that such words in fact take the form /daei, deuule, daee/; the nature of the relevant contrasts is controversial, but we have to take a stand, since these sound sequences are frequent. Glides in pre-vocalic position are both (on the account that postulates them) rare and controversial; we can afford to avoid taking a stand on them; in this dissertation, we choose to do exactly this.

Of the nasals *m n ng,* velar *ng* is always post-vocalic. The place of articulation of *n* partly assimilates to the following sound. This assimilation also affects the lateral liquid *l* and, to an uninvestigated extent, the median (i.e. non-lateral) liquid *r*. After a dental, *r* tends to be a tap or trill. Elsewhere standard speakers tend to use an undescribed allophone – some sort of frictionless continuant, to use the terminology of the International Phonetic Association.

Unaffected by assimilation are the never-initial median flaps *R* and *Rh* which start out with 'r-colouring' but for careful speakers include a momentary stroke of the curled up tongue-tip against the post-alveolar region. Final *Rh* deaspirates regularly and non-final *Rh* variably. Eastern Bangla influence has tended to merge *r, R, Rh* into *r* since 1947. Conservative speech distinguishes *garo, gaRo, gaRho* 'Garo (mountains), place (your knees) on the ground, dense'.

Many standard Bangla speakers distinguish alveolar *s* from palato-alveolar *S*. The pronunciation of *S* varies: apico-post-alveolar (especially for upper-middle-class female

speakers), lamino-alveolar with palatalization, lamino-post-alveolar, or (rarely) alveolo-palatal.

The affricates *c ch j jh* become *s s z z* before certain consonants; this detail is ignored in the transcription here, as it is in written Bangla. Retroflex *T Th D Dh* differ from dental *t th d dh* as to the part of the tongue involved. Thus, English *t d* sounds are perceived by Bangla-speaking listeners as half-way or entirely retroflex when apical and as half-way or entirely dental when laminal. Labial *p ph b bh* and velar *k kh g gh* require no comment.

The voicing of *h* varies. The determinants and nature of its voicing are controversial. Its contextual determinants in Bangla are unknown.

Bangla vowels undergo vowel harmony. The most important rule is one whereby a following high vowel or glide (semi-vowel) raises a preceding vowel or glide from low to mid or from mid to high. There are also systematic alternations involving vowels and glides. As labels for morphemes which (owing to the processes mentioned) have alternate phonological shapes, this volume will use their underlying phonological form. Ordinary citation forms use the transcription which is phonemic for structuralists and roughly phonetic for Chomsky and Halle.

2.2 Declension

In Bangla, adjectives do not inflect, either in agreement or for comparison. Declension only affects NP. For ease of exposition, we will first treat pronouns with nouns as if they both belong to the N category. Chapter 3 argues, however, that some pronouns are in fact not nouns.

2.2.1 Personal pronouns

GLOSS	NOM SG	NOM PL	OBJ SG	OBJ PL	GEN SG	GEN PL
I/we	ami	amra	amake, amaY	amader	amar	amader
you$_1$	tuy	tora	toke	toder	tor	toder
you$_2$	tumi	tomra	tomake, tomaY	tomader	tomar	tomader
you$_3$	apni	apnara	apnake	apnader	apnar	apnader

The ‘you’ forms are arranged in ascending order of honorificity, from intimate ‘you$_1$’ to honorific ‘you$_3$’. Notice the *u/o* alternation for the intimate and neutral forms which derive from /toi tomi/ via the phonological processes alluded to in section 2.1. Notice the absence of locatives. It turns out that pronouns which denote persons invariably lack the locative. If we make this our key criterion for personal-pronoun-hood and assume a broadminded criterion for pronounhood, then the personal pronouns not included in the above chart are:

GLOSS	NOM SG	NOM PL	OBJ SG	OBJ PL	GEN SG	GEN PL
they$_1$	e	era	eke	eder	er	eder
they$_2$	ini	eMra	eMke	eMder	eMr	eMder
they$_3$	o	ora	oke	oder	or	oder
they$_4$	uni	oMra	oMke	oMder	oMr	oMder
they$_5$	Se	tara	take	tader	tar	tader

they$_6$	tini	taMra	taMke	taMder	taMr	taMder
themselves	nije	nijera	nijeke	nijeder	nijer	nijeder
others	onne	onnera	onnoke	onnoder	onner	onnoder
everyone	SObay	----	SObayke	----	SObar	----
someone	kew	----	kawke	----	karur	----
who?	ke	kara	kake	kader	kar	kader
who$_1$	je	jara	jake	jader	jar	jader
who$_2$	jini	jaMra	jaMke	jaMder	jaMr	jaMder

To keep the glosses short, only the plural forms have been given, which is not too inappropriate, since modern English usage countenances *they, them,* etc. for a singular referent of unknown or irrelevant gender. 'They$_1$' and 'they$_2$' are Proximal forms, 'they$_3$' and 'they$_4$' Distal forms, 'they$_5$' and 'they$_6$' Sequent forms; Distals and Proximals are Demonstratives. 'Who?' and 'who' are Interrogative and Relative respectively.

It is difficult and perhaps unreasonable to pick out a subset of the chart and call it 'the set of Third Person Pronouns'. Chatterji 1968 cites the 'they$_5$' and 'they$_6$' forms as specifically Third Person forms. However, if one wishes to retain some cross-linguistic semantic significance for the term 'Third Person', one must be able to say that Eastern Bangla speakers tend to use the Sequent forms as Third Person pronouns and to reserve the Distal forms for a clearly pointing, Demonstrative function, whereas Western Bengalis use Distals as Third Person pronouns and to reserve the Sequent forms for a distinctly discourse-related cross-referential function. So, let us propose that the term Third Person Pronoun refer to the purposes served in English by *he, she, they* as opposed to *this person, that person, these people, those people*. For formal grammatical description, then, we shall call Bangla *Se* and *tini* 'Sequent' pronouns in contrast to the Demonstrative pronouns *e, ini, o, uni*. On a rule of thumb basis, it is convenient to consider that Proximals refer 'in my or our orbit', Distals 'in your orbit', and Sequents 'in some other orbit'. But the last word on the relation between Demonstratives and Sequents – a relation which straddles the fence between anaphora and deixis – cannot be said in advance of detailed research on Proximal /Distal /Sequent systems in Bangla and other languages that have them (Spanish, Latin, Japanese, Sanskrit).

One might object that by implying that the two charts above exhaust the class of Personal Pronouns I am failing to express the traditional notion of Personal Pronouns as a class including inanimates. But the concept that these charts do demarcate – that of a class of words which decline, in Bangla, for the Nominative, Objective, and Genitive, but not for the Locative – may prove more useful than the traditional notion. One example of such usefulness follows.

The set of English possessive pronouns includes *mine, ours, yours, his, hers,* and animate *theirs*, but for many speakers excludes *its* and inanimate *theirs*. Thus, the well-formed *We harm our neighbours and you harm yours* contrasts with the ill-formed **We harm our neighbours and the factory harms its*. And *The plank can support our weight but not theirs* has

no reading where *theirs* means, say, 'of the bricks'. If the universal definition of Personal Pronoun imposes a requirement of personhood, this otherwise puzzling fact about English falls out as a routine consequence, as does the prima facie puzzling fact that Bangla Personal Pronouns lack a locative. It so happens that person-referring Nouns in Bangla have very reluctant, rare locative singulars (serving non-locative purposes, for the most part) and no locative plurals at all; and Personal Pronouns carry this lacuna to its logical conclusion. (Some varieties of English apparently do not exclude inanimate possessive pronouns; my argument is based on those that do.)

Let us, on these grounds, pledge tentative allegiance to a universal definition of Personal Pronoun which requires such Pronouns to refer to Persons. The other half of the requirement – that Personal Pronouns be pronouns – is more difficult to formulate. No one quite knows what pronouns are any more. An attempt to characterize the notion of Pronoun follows. We then move on to Non-Personal Pronouns.

Via cliticization, incorporation, and concord, Pronouns manage to set up a close relationship with the inflectional morphology of verbs in most languages. In a language (like Japanese or Bahasa Indonesia) where this takes place only marginally or not at all, one tends to find independent grounds for doubting the wisdom of postulating a class of Pronouns for the grammar of that language. Let us take the pronoun-verb nexus seriously and ask how the structure of the class of pronouns mirrors, if it does, their role with respect to the conjugation.

A rough answer runs as follows. Verbs inflect for some of the Dimensions (Dimension is the superordinate term Chomsky (1965) proposed for Person, Number, and the like) that pronouns inflect for. On these dimensions, finite verbs always adjust their inflection to that of a properly related pronoun (in most languages, in most contexts, a subject pronoun), if overtly present, or even if present in spirit. The inflectional behaviour of verbs helps identify some dimensions (Person and Honorificity, in Bangla) that partition the class of Pronouns into subclasses. These subclassifications tend to impose rigid demarcations. Thus, if First vs non-First Person is a relevant contrast, then any given Pronoun will be one or the other. (Cf. the mass/count distinction in English as an example of a non-rigid demarcation. The noun *sugar* can be used as a mass or as a count noun, although *intelligibility* is only a mass noun and *city* only a count noun.) Given more than one relevant dimension, the rigid subclassification system freezes Pronouns into paradigms of the familiar sort, presented in traditional handbooks in chart format.

To see how this makes pronouns any different from regular nouns, consider such facts as that, while 'Ashish knows' can be said either with an honorific verb (*aSiS janen*) or a non-honorific verb (*aSiS jane)* depending on presuppositions about the person called Ashish, *Se jane* and *tini janen* are obligatory for the non-honorific and honorific versions of '(s)he knows'. It would be ungrammatical to say *Se janen or tini jane.* Contrast these impossible forms with the pragmatically odd forms *prodhanmontri jane* 'the Prime Minister knows (non-honorific)' and *aSiSer Soddojato meYe janen* 'Ashish's new-born daughter knows (honorific)', which become non-odd if the prime minister's uncle is speaking or if Ashish's new-born daughter is the heroine in a biography where we know that she is really a big shot in the making and deserves honour each time she is referred to, at any age. Even such extreme contexts do not spare **Se janen, *tini jane*. *Se* and *tini*, being Pronouns, irrevocably belong to paradigms which assign them non-honour and honour respectively. Nouns which are not Pronouns get their Honour feature value from the context or from standardized contexts and are thus always more or less mutable with respect to this feature.

Whether or not Pronouns are all Nouns (chapter 3 will show that they are not), Pronouns do belong to paradigms that assign rigid, immutable features along paradigmatic Dimensions such as Honour and Person. Extending ordinary usage slightly, let us say Pronouns 'inflect' along Dimensions like Person and Honorificity, just as verbs do. More accurately, we can say that *The Personal Pronoun* inflects for Person, Honour, and other relevant Dimensions.[4] Equating the completeness and rigidity of paradigms with morphological inflection – postulating inflection wherever one finds rigid paradigms – is good methodology because a sharp line is thus drawn between derivational morphology (with its fragmentary paradigmlike formations) and inflectional morphology (which is all paradigms). The idea of drawing this sharp line stems from Aronoff (1976).

Once we decide to define the class of pronouns in terms of inflectional Dimensions in paradigms, we have to set up some Dimensions which pertain only to pronouns (just as there are Dimensions like Tense which pertain only to Verbs). Looking back at the chart of Personal Pronouns we find a clear case of such a Dimension: the one along which words beginning with /k/ vary. Interrogative *ke* and indefinite *kew* are related in that the suffix *w* (alternating with *W, o, u, ru*) added to any interrogative pronominal form yields an indefinite form corresponding to it. In a study of this suffix (Dasgupta 1979) I have shown that it is best regarded as an inflectional suffix. Its presence and absence, then, are two points on an inflectional Dimension.

One purely pronominal Dimension we have already set up, in effect, is that of Distance, on which the points Proximal and Distal lie. Only Demonstrative forms inflect for that dimension as it stands. The Distance dimension is so strongly reminiscent of the First/Second Person contrast that one is tempted to flag the forms in the small chart (at the beginning of this section) with a special label like 'Non-Phoric Personal Pronouns' and then to state that Demonstratives and Non-Phorics inflect for Distance, as follows:

	Non-Phoric	Demonstrative
Distal	Second Person	'Yonder person (in your orbit)'
Proximal	First Person	'This person (in my orbit)'

Such a move would make for efficiency and compactness – as did Jakobson, Fant, and Halle's (1952) proposal that retroflex sounds and emphatic (velarized/pharyngealized) sounds both be regarded as 'flat', since no language distinguishes them. The community of phonologists greeted the 'flatness' proposal with uncritical and silent rejection. No doubt a similar fate awaits my proposal. But I will make it anyway. For simplicity's sake I will keep

[4]The reason that this statement is 'more accurate' is as follows. When we say verbs inflect for, say, Person, we can give examples of many verbs doing it: *achi* 'am', *ache* 'is'; *bachi* 'choose', *bache* 'chooses'. But when we say 'pronouns inflect for Person', we can cite *tumi, tomra, ami, amra* corresponding to *achi, ache, bachi, bache*, but it does not have the same effectiveness as an example. The difference between the verbs *ach* and *bach* is a true lexical difference. In contrast, the difference between 'the pronouns *am-* and *tum-*' is itself a fact about the inflectional system: *am-* and *tum-* represent two points on a Dimension.

saying ‘First Person’ and ‘Second Person’, but I will assume that they are really points on the grid given above.

These and many other dimensions conspire to form multiplex paradigms to which pronouns rigidly adhere, with or without corresponding inflectional behaviour in verbs. How does this help answer our original question, which was ‘What is a Pronoun?’?

One way to define ‘Pronoun’ is to construct a formal object which one calls the class of pronouns and then to call its members pronouns. The above discussion has been the beginning of such a constructive definition. I cannot hope to complete the project here, but only to offer some conjectures about how it might or should be continued, and what one might expect from a complete system of interlocking Dimensions along which Pronouns inflect.

What makes the project possible is the high degree of regularity in the system of Pronouns – what I have called their rigid adherence to paradigms – and the complexity of this regularity – what one might express by saying that the system of pronouns hangs together as a system of subsystems, each subsystem having its own ‘local’ laws. Concretely, this hanging together means that some Dimensions presuppose others and that some Dimensions intersect. For example, Distance and Honour intersect, and Distance presupposes Demonstrativity or non-phoricness. It seems fair to say that, if a Dimension is relevant to Pronouns, it either presupposes or intersects with at least one other Dimension which is relevant to Pronouns. Thus, the structure of the paradigmatic systems of Pronouns lends itself to inquiry. Each Dimension, as one begins to understand it, leads to other, related Dimensions.

Decisions as to whether word set X belongs to the class of pronouns will, as the project progresses, tend to hinge on the compactness and efficiency of the set of Dimensions required to accommodate X versus the compactness of the set of Dimensions required to exclude X. One case in point is the set comprising *nije* ‘themselves’ and *onne* ‘others’. In this thesis I take no position on whether these words are pronouns. But a serious study of the question will have to ask and answer, at many levels of analysis, the question ‘Is *nije* opposed to *onne* as Proximal to Distal?’

Suppose that the project of research envisaged is completed and that, at least for Bangla, we have obtained a fully worked out grid of grids for pronouns. In this imaginary situation, how are we better off than we were? What properties can one expect the grid of grids to have?

It seems likely that it will have at least one property of theoretical interest: the property of defining each pronoun as grammatically nothing but a set of values for various pronoun-relevant Dimensions. Members of major word classes have more content than that. *Petty* and *tacky*, for examples, are two adjectives with probably the same grammatical features. But, if my hunch is right, there can be no pronoun pair like the pair *petty – tacky;* pronouns with the same features will be the same pronoun. (Dialect mixture will keep producing marginal exceptions, e.g. *himself vs. hisself*.) If this result is obtained, it will lend credence and content to the widespread feeling that pronouns constitute a non-lexical category, that they are ‘function words’, not ‘content words’.

2.2.2 Non-personal pronouns

To begin with the shibboleth: non-personal pronouns in Bangla have locative forms. And they have number-invariant indefinite forms as well as number-inflected definite forms:

INDEFINITE:[5]				
GLOSS	NOMINATIVE	OBJECTIVE	GENITIVE	LOCATIVE
this	e	eke	er	ete
that$_1$	o	oke	or	ote
that$_2$	ta	take	tar	tate
all	SOb (-kichu)	SOb-kichuke	SOb-kichur	SOb-kichute
something	kichu	kichuke	kichur	kichute
what?	ki	kake	kiSer	kiSe
what	jar	jake	jar	jate
DEFINITE				
this SG	eTa	eTake	eTar	eTate
this PL	Egulo	eguloke	egulor	egulote
that$_1$ SG	oTa	oTake	oTar	oTate
that$_1$ PL	Ogulo	oguloke	ogulor	ogulote
that$_2$ SG	SeTa	SeTake	SeTar	SeTate
that$_2$ PL	Segulo	Seguloke	Segulor	Segulote
all SG	SObTa	SObTake	SObTar	SObTate
all PL	SObgulo	SObguloke	SObgulor	SObgulote
some one	konoTa	konoTake	konoTar	konoTate
which one?	konTa	konTake	konTar	konTate
which ones?	kon$gulo	kon$guloke	kon$gulor	kon$gulote
which one	jeTa	jeTake	jeTar	jeTate
which ones	jegulo	jeguloke	jegulor	jegulote

With some semantic effect E and under some set of conditions C, both of which are to be determined by future research, the *e o Se* stems in the definite part of the chart optionally enlarge to become the Augmented forms *ey oy Sey*. Set C includes the condition 'before *Ta* or *gulo*'. In casual speech, *eyTa oyTa SeyTa* often vowel-harmonize to *eyTe oyTe SeyTe*, while *eygulo* etc. remain intact since *gu* already has a high vowel.

In slightly less casual speech, the single suffix *Ta* gives way to a pair of contrasting suffixes, *Ti* for small, respectworthy or feminine referents and *Ta* for the opposite thereof. Correspondingly, *gulo* gives way to a *guli* versus *gulo* contrast along the same lines. In still more formal speech and in much writing, *Ti* and *guli* completely supplant *Ta* and *gulo* for all referents. Although this study draws its examples consistently from casual speech, the special considerations that follow require us to pay attention to the stylistic *Ta-Ti* alternation.

2.2.3 More about *Ta*: Numerals and Denominators

Sequences like *Ek peYala ca* "one cup tea" 'a cup of tea', *tin gOj Suto* "three yard

[5] In the 'which ones' line, we show a syllable boundary between /n/ and /g/ to ensure that the <ng> sequence is not read as a velar nasal – a small price to pay for the use of this convenient digraph. [A 2020 footnote.]

thread" 'three yards of thread', *du koTi lok* "two crore person" 'two crores (20,000,000) of people', etc. are usually described as Numeral + Measure Word + Noun. In contrast, sequences like *Ek jon lok* "one fellow person" 'a person', *tin khana boy* "three piece book" 'three books', *paMc Ta boRi* "five item pill" 'five pills', etc. are usually described as Numeral + Classifier + Noun. However, the difference between alleged Measure Words and alleged Classifiers boils down to a difference between the English glosses called for: Measure words translate easily, while Classifiers do not, except in Pidgin English, where *one fellah man* would raise no eyebrows. Let us ignore this pointless distinction and call both Classifiers and Measure Words Denominators and postulate phrase structure rules to generate them. Rule (1) expands Noun Phrase into optional Nume-prime plus obligatory Noun. In current usage, prime notation and bar notation are interchangeable; Nume-prime (NUM′) could have been written Nume-bar (with a bar above NUM) instead. By writing NUM′ instead of NUML′ (Numeral-prime) or DENOM′ (Denominator-prime) I am leaving it up to the reader, or future research, to decide whether it is NUML or DENOM that is the head of NUM′; the decision is slightly difficult and need not detain us. If more accurate studies of the NUM′ system show that rule (2) should read NUM′ → NUML (DENOM), with the Numeral node obligatory and the Denominator optional, then NUM′ should read NUML′. My decision to keep this problem on hold is encoded in the formal gap between NUML and NUM.

(1) NP → (NUM′) N
(2) NUM′ → NUML DENOM

If both (1) and (2) apply, the resulting sequences have the form NUML DENOM N, as in (3). There are grammatically related sequences of the form DENOM NUML /ek/ N, as in (4).

(3) a. du peYala ca "two cup tea" 'two cups of tea'
b. tin gOj Suto "three yard thread" 'three yards of thread'
c. car koTi lok "four crore person" '40,000,000 people'
d. car jon lok "four fellah person" 'four people'
e. tin khana boy "three piece book" 'three books'
f. paMc Ta boRi "five item pill" 'five pills'

(4) a. peYala duek ca "cup two-*ek* tea" 'a cup or two of tea'
b. gOj tinek Suto "yard three-*ek* thread" 'about three yards of thread'
c. koTi carek lok "crore four-*ek* person" 'some 40,000,000 people'
d. jOna carek lok "fellah four-*ek* person" 'roughly four people'
e. khan tinek boy "piece three-*ek* book" 'three books or so'
f. goTa paMcek boRi "item five-*ek* pill" 'about five pills'

I assume that one or more rules relate (3) to (4). Thus, I equate the Denominator elements in (3) with those in (4). It is not difficult to equate *peYala, gOj,* and *koTi* with *peYala, gOj,* and *koTi,* and only a bit harder to group *jon* and *jOna* into one morpheme and *khana* and *khan* into another. Some readers may balk at a morpheme whose allomorphs are *goTa* and *Ta.* To them it may be pointed out that a stylistic alternation between *Ta* and *Ti* corresponds to one between *goTa* and (underlying /goTi/>) *guTi,* and that there is no argument against uniting *goTa* and *Ta*

into one morpheme.

What syntactic category does this morpheme belong to? (3) and (4) indicate that the morpheme *(goTa)* is a DENOM. But in the charts of sections 2.2.1 and 2.2.2 it is the pronominal inflectional suffixes *gulo* and *ra* that *Ta* patterns with. Thus we find *eTa* 'this' alongside *egulo* 'these things' and *era* 'these people', but no **e koTi lok* "this crore person": (*go)Ta* alone among all the Denominators seems to share the distribution of *gulo* and *ra*. So, we must assume, for the time being, that *(go)Ta* is both a DENOM and a declensional suffix. Section 2.2.6 will offer a wider range of data and show that *(go)Ta* is always a Denominator.

2.2.4 Human nouns

Turning to the declension of nouns other than pronouns, we look first at Human Nouns. These and only these co-occur with the Denominator *jon*. Some of them also co-occur with the default non-human Denominator *Ta*. Human Nouns also exhibit, though less rigidly, two features of Personal Pronouns: the lack of a Locative and the obligation to employ an overt pluralizer whenever the referent is, in Dougherty's (1970) sense, semantically non-singular. Human Nouns exhibit these features less rigidly in that a rare and archaic usage requires singular locative forms of human nouns (although plural locative forms are so thoroughly non-existent that their shapes are unimaginable) and in that some existential and negative contexts countenance a human noun with a semantically non-singular referent lacking a tangible pluralizer.

Since Personal Pronouns cannot be numbered but Human Nouns can (cf. **tin jon amra* 'three we' and **tin jon ami* 'three I', but well-formed *tin jon amla* 'three bureaucrats'), the latter have a wider range of overt pluralizers to choose from. A Personal Pronoun with a semantically non-singular referent must end in *ra* or *der*. A Human Noun with such a referent may either end in *ra* or *der* or end in *gulo/guloke/gulor* or be in construction with a word indicating plurality, as in *tin jon kormi* 'three employees'.

The paradigm for *chele* 'boy, son' is as follows.

Subparadigm	NUMBER	NOMINATIVE	OBJECTIVE	GENITIVE
ra-jon	SG	chele	cheleke	cheler
ra-jon	PL	chelera	cheleder	cheleder
ra-jon	numerated	tin jon chele	tin jon cheleke	tin jon cheler
gulo-Ta	SG	cheleTa	cheleTake	cheleTar
gulo-Ta	PL	chelegulo	cheleguloke	chelegulor
gulo-Ta	numerated	tin Te chele	tin Te cheleke	tin Te cheler

The *gulo-Ta* SG and PL forms are definite: *cheleTa* 'the boy', *chelegulo* 'the boys'.

Human Nouns which numerate with both *jon* and *Ta* have the complete paradigm given above. Human Nouns like *montri* 'minister' or *kormi* 'employee' which numerate only with *jon* and not with *Ta* have only *ra-jon* forms. As the glosses for *cheleTa* and *chelegulo* indicate, this means that, for nouns belonging to the *kormi* subclass (rather than the *chele* subclass) of Human Nouns, it is impossible to say 'the N' in Bangla – quite impossible in the singular and impossible to do unambiguously in the plural. (The qualification 'unambiguously' refers to the fact that *ra* forms sometimes take on a 'definite' colouring through an interaction of context

with the semantics of plurality.) In practice, speakers (and especially authors) often solve this problem by resorting to formal style and using *Ti* rather than *Ta*: *kormiTi* 'the employee'. But this always fails in the plural, since *guli* carries an un-overridable connotation (a connotation mid-way between non-human and extremely non-honorific), and sometimes fails even in the singular, if the noun has an especially honourable referent: *prodhanmontri* 'Prime Minister', **prodhanmontriTi*. The language has patch-up devices for such emergencies.

The above reads like a sacrilegious resort to translation-based thinking – setting Definiteness up as an absolute and asking how Bengalis go about expressing it. But the translation, if any, is not from other languages to Bangla, but from the part of Bangla that distinguishes indefinite and definite (*chelera*, 'boys' and *chelegulo* 'the boys') to the part which doesn't (*kormira* 'employees', nothing for 'the employees'). This is like lamenting the non-occurrence of Locatives in the paradigms of Personal Pronouns and in the non-singular parts of the Human Noun declension – quite legitimate, to my mind, within a methodology that does not depend on translation across languages for its essential moves.

2.2.5 Nonhuman nouns

Nonhuman nouns may numerate with *Ta* (they all have this option) and *khana* (some of them have this option). Historically *Ta* and *khana* come from words that meant 'whole' and 'piece', respectively. This 'unwhole' nuance still clings to *khana*, making it impossible to say *Ek khana SOhor, Ek khana patihaMS, Ek khana belehaMS* 'a town, a duck, a goose', though one may say *Ek khana megh, Ek khana tOkta* 'a cloud, a plank/board'. *Ek khana kagoj* 'a paper' requires *kagoj* to mean '(single) newspaper' or 'piece of paper' and excludes any abstract reading. The paradigm for *kagoj* is as follows.

SUBPARADIGM: gulo-Ta				
NUMBER	NOMINATIVE	OBJECTIVE	GENITIVE	LOCATIVE
SG	kagojTa	kagojTake	kagojTar	kagojTate
PL	kagojgulo	kagojguloke	kagojgulor	kagojgulote
Numerated	tin Te kagoj	tin Te kagojke	tin Te kagojer	tin Te kagoje
SUBPARADIGM: khana				
SG	kagojkhana	--	kagojkhanar	kagojkhanate
Numerated	tin khana kagoj	--	tin khana kagojer	tin khana kagoje

2.2.6 More on definiteness

The traditional view that the Definiteness of *Ta* and *gulo* forms is due to a definitizing function of *Ta* and *gulo* seems to me somewhat suspect. Let us examine the relation between (5) and (6) below. Readers familiar with Bangla spelling will notice that I am violating the rule that *Ta* and *khana* are to be attached to their immediate left neighbour.

(5) boy du To "book two item" 'the two books'
boy tin Te "book three item" 'the three books'

boy paMc Ta "book five item" 'the five books'
ciThi du khana "letter two piece" 'the two letters'

(6) du To boy "two item book" 'two books'
tin Te boy "three item book" 'three books'
paMc Ta boy "five item book" 'five books'
du khana ciThi "two piece letter" 'two letters'

The NUML DENOM sequences at (6) derive from rules (1) and (2). The N NUML DENOM sequences (5) derive either from rules (7) and (2) or from an application to (6) of a syntactic movement rule like (8).

(7) NP → N (NUM′)
(8) NUM′ Postposing
S.D.: NUM′ – N
S.C.: 1 2 → 0, 2+1

Whichever solution is chosen, the correct analysis of (5) must on the one hand provide for Definite semantic interpretation and on the other hand respect the limits of phenomenon (5). These limits include (i) the non-occurrence of medium or large numbers (**ciThi dOS khana* 'the ten letters', *boy aT Ta* 'the eight books') and (ii) the fact that the DENOM must be *Ta* or *khana* (thus, no **chele tin jon* 'the three boys' corresponding to *tin jon chele* 'three boys'). These restrictions can be built into the interpretation rule which is to yield the appropriate Definite reading. So, the restrictions do not directly argue in favour of either (7) or (8), if one assumes that semantic interpretation may – or does – accept the output of (8) as input.

However, if one assumes instead that the only input to (the relevant submodule of) semantic interpretation is 'deep' (base-generated, pre-transformational) structure, then one must choose between (7) and some version of (8). That version of (8) will have to be complicated. If (7) does not exist and the ±Definite feature must be read off of deep structure, then the deep structure of (5), unlike that of (6), will need to carry a +Def feature to trigger both the semantic rule giving the Definite reading and the transformation, (8), which gives the surface structure for that reading. So, the formulation of (8) will need to refer to that feature. If one assumes deep structure semantics, it will also follow that the limits of phenomenon (5) are best stated as lexical selectional features. Thus, one may mark the small numbers as admissible in a +Def construction whose DENOM is *Ta* or *khana*. Or, if one does not want a transformational analysis and hence does not posit +Def, one can have a phrase structure rule (7) and make additional lexical assumptions that have the effect of making only small numbers and only the Denominators *Ta* or *khana* admissible in the structure generated by (7).

Although the assumption that (relevant) semantic rules apply to 'shallow' structure (the output of movement rules) and the assumption that they apply to deep structure both lead to a situation where it is difficult to choose between the base analysis (7) and the transformational analysis (8), there is a slight difference, as shown above. Namely, the deep semantics assumption, unlike the shallow semantics assumption, forces us to complicate (8). This may be a point in favour of the shallow semantics assumption. But the (7) vs. (8) issue remains unresolved, so the argument is inconclusive.

Setting these questions aside, let us turn to (9) and (10).

(9) boy Ta “book item” ‘the book’
*boy Ek Ta “book one item”
boy khana “book piece” ‘the book’
*boy Ek khana “book one piece”

(10) Ek Ta boy “one item book” ‘a book’
Ek khana boy “one piece book” ‘a book’

It seems reasonable to suppose that whatever rule relates (5) to (6) must also relate (9) to (10). To formalize this correspondence we need to integrate (9) into (5) via a rule of *Ek*-Deletion which reduces *boy Ek Ta* to *boy Ta* and *boy Ek khana* to *boy khana*. This brings us back to the opening sentence of this section. The forms *boyTa* and *boykhana* ‘the book’, traditionally assumed to be inflectional forms of the word *boy* ‘book’, now appear to be the product of at least one transformation – of two, if one accepts (8) rather than (7).

The other ‘definite’ form, N-*gulo*, may also reflect a deletion rule. If one groups (11) with (5) and (12) with (6), then a rule of *kOtok*-Deletion would derive *boy gulo* from *boy kOtok gulo*.

(11) boy gulo “book GULO” ‘the books’
*boy kOtok gulo “book some GULO”

(12) kOtok gulo boy “some GULO book” ‘some books’

Here is one argument in support of *Ek*-Deletion and *kOtok*-Deletion. Bona fide declensional suffixes typically do phonological damage. The pluralizers *ra* and *der* require or allow special allomorphs of the stem they are suffixed to. Thus *lok* ‘person’ becomes *loke* in *loke-ra, -der*. But with *Ta* and *gulo* you just get *lok-Ta, -gulo*. Again, the Personal Pronoun *Se* plus *ra* and *der* yields *tara* ‘they’, *tader* ‘them /their’. But *Se* remains *Se* when *Ta* or *gulo* is suffixed. The Case suffixes *ke* and *r* also select allomorphs of their stems complexly (Ferguson 1945: 75-78); not so *Ta* or *gulo*. The utter lack of phonological interaction of *Ta* and *gulo* with the preceding stem contrasts with the behaviour of bona fide declensional suffixes.

This argument can be rephrased as an appeal to the simplicity of linguistic descriptions. It is simpler to be able to say ‘the Pronoun *Se* has the allomorphic shape *ta* whenever it takes a declensional suffix’ than to have to say ‘the Pronoun *Se* has the shape *ta* whenever it takes a declensional suffix other than *Ta* or *gulo*’. Let us therefore drop the assumption that *Ta* and *gulo* are declensional suffixes.

The sketch presented here will be revised slightly in chapter 3.

2.3 Conjugation

Most of the complicated word phonology that goes on in Bangla is in Verb words. Although this is not a treatise on phonology, it seems only fair to present some analysis of the phonological phenomena which the reader will, in any case, notice in the data. I do not attempt a defence of the analysis offered, although I believe a defence is possible.

2.3.1 Simple (aspect-free) forms

A simple finite verb form consists of stem plus optional tense plus optional finite ending. Here is a paradigm for the stem *kOr* 'do' given in phonetic representation ('systematic phonetic' for Chomsky & Halle, 'phonemic' for Ray et al.) A dash marks a non-occurring combination. Notice the processes of vowel harmony and glide deletion. A 'notional' paradigm is given below the 'formal' paradigm.

ENDING	TENSE zero	TENSE /yl/	TENSE /yb/	TENSE /yt/
zero	kOr	--	--	--
/am/	--	korlam	korbam	kortam
/i/	kori	korli	korbi	korti
/iS/	koriS	--	--	kortiS
/uk/	koruk	--	--	--
/un/	korun	--	--	--
/e/	kOre	korle	korbe	korte
/en/	kOren	korlen	korben	korten
/o/	kOro	korlo	korbo	korto
/yo/	koro	--	--	--

TENSE	3 PERSON neutral	NON-1 PERSON honorific	2 PERSON neutral	2 PERSON intimate	1 PERSON
Indic. Pres.	kOre	kOren	kOro	koriS	kori
Imp. Pres.	koruk	korun	kOro	kOr	kori
Past	korlo	korlen	korle	korli	korlam
Indic. Fut.	korbe	korben	korbe	korbi	korbo
Imp. Fut.	korbe	korben	koro	koriS	--
Conditional	korto	korten	korte	korti	kortam

Some Eastern dialects have *korbam* instead of *korbo*. Some Western dialects have *kortiS* instead of *korti*. While *korbam* is not considered standard, *kortiS* is.

A simple non-finite verb form consists of stem plus non-finite ending (/ye/ or /yte/ or /yle/). Thus, the stem /*kOr/* yields the forms *kore, korte, korle.* One might gloss these forms, which are, respectively, Conjunctive, Infinitive, and Protactic, as 'doing' or 'having done', 'to do', and 'if …does/did'.

A simple nominal verb form consists of stem plus ending (/yba/ or /Wa/~/no/). The stem /*kOr/* yields the forms *korba, kOra*. One might gloss these (respectively Oblique Gerund and Gerund-Participle) forms as 'doing' and as 'doing/done'. Although the mother node of these forms is perhaps never V (chapter 4 discusses this point for Gerund-Participles), Oblique Gerunds and Gerund-Participles belong to the inflectional system of the verb. First, they are formed from verb stems. Second, every verb stem that has a finite conjugation displays these forms as well, so that one cannot call /yba/ and /Wa/~/no/ derivational suffixes.

The word level rules of phonology informally indicated below derive the systematic phonetic representation of verb forms from their underlying phonological representation. Since we are not concerned with formalism here, this formulation avoids official notation and features. In the following statements, V stands for Vowel, G for Glide (semi-vowel), K for non-glide consonant, C for glide or consonant, V* for a vowel no lower than its predecessor, and G** for a glide homorganic to the vowel preceding it. Rule (13:11) comes into play only in section 2.3.2, but for convenience appears here with its fellow rules. A more careful analysis would try to explain the inapplicability of some of these rules to non-verbs, and would distinguish morphology and phonetics from phonology proper that this formulation mixes them up with. The numbering encodes extrinsic ordering in obvious ways: 5a and 5b, for instance, are not crucially ordered relative to each other, but they have to follow 4 and precede 6.

(13) 1 /yb/-Truncation: y → Ø / V__

2a Metathesis: Cy → yC

2b Lowering: {i, u} → {e, o} /__C_0 <Wa>]$_{verb}$: <complexly optional>

3 /a/-Fronting: a → E/ __ <#(C)>__y<C>V

4 Non-initial Raising: {E, O} → {e, o} / VC_0__

5a Regressive Raising: {E, O, Y, W, e, o} → {e, o, y, w, i, u}/__C_0{i, y, u, w}

5b Progressive Raising: a → o /{i, u, y, w}C_0__]$_{verb}$

6 Preconsonantal Front Glide Drop: {y, Y} → Ø / __K_1V

7 /iy/ -Weakening: i → y / {G__yV, CC__ye}

8 Postconsonantal Onglide Drop: G → Ø / K__

9 /e/-Lowering: e → E / __{e, o} (nonstandard: __ {e, o, W})

10 /E/-Backing: E → a / __o

11 /cch/-Simplification: c → Ø / C__ch

12 /h/-Drop: h → Ø / {V, G} __

13 Last Vowel Weakening: V* → G/ V__{C, #}

14a Pre-coda Glide Drop: G → Ø / V__G_0C{C, #}

14b Intervocalic Glide Drop: G** → Ø / V__V

Some derivations illustrating the use of these rules follow. The notation ‘ABCD n XYZW’, where ABCD and XYZW are strings and n is one of the rule numbers from 1 through 14b, means that ABCD becomes XYZW by application of rule n of (13). When n = 0, no rule has applied. The symbols hyphen (-) and space ( ) indicate morpheme boundary and word boundary respectively. As the derivations illustrate, the numbering of rules corresponds to crucial ordering relations. Thus 13 must precede 14a, which need not precede or follow 14b.

Stem: /kOr/ ‘do’

kOr-i 5a kor-i
kOr-uk 5a kor-uk
kOr-yo 2a kOyr-o 5a koyr-o 6 kor-o
kOr-yto 2a kOyr-to 5a koyr-to 6 kor-to
kOr-Wa 8 kOr-a

Stem: /dEkh/ ‘see’

dEkh-i 5a dekh-i
dEkh-uk 5a dekh-uk
dEkh-yo 2a dEykh-o 5a deykh-o 6 dekh-o
dEkh-yto 2a dEykh-to 5a deykh-to 6 dekh-to
dEkh-Wa 8 dEkh-a

Stem: /tul/ ‘lift’

tul-i 2b tol-i 5a tul-i
tul-o 2b tol-o
tul-yo 2ab toyl-o 5a tuyl-o 6 tul-o
tul-yto 2ab toyl-to 5a tuyl-to 6 tul-to
tul-Wa 2b tul-Wa 8 tol-a

Stem: /kin/ ‘buy’

kin-i 2b ken-i 5a kin-i
kin-o 2b ken-o
kin-yo 2ab keyn-o 5a kiyn-o 6 kin-o
kin-yto 2ab keyn-to 5a kiyn-to 6 kin-to
kin-Wa 2b ken-Wa 8 ken-a

Stem: /hO/ ‘be, become’

hO-i 5a ho-i 13 ho-y
hO-uk 5a ho-uk 13 ho-wk 14a ho-k
hO-yo 5a ho-yo
hO-yto 5a ho-yto 6 ho-to
hO-ybe 1 hO-be

hO-ybi 1 hO-bi 5a ho-bi
hO-Wa 0 hO-Wa

Stem: /Su/ ‘lie down’

Su-i 2b So-i 5a Su-i 13 Su-y
Su-o 2b So-o 13 So-W
Su-yo 2b So-yo 5a Su-yo
Su-uk 2b So-uk 5a Su-uk 13 Su-wk 14a Su-k
Su-yto 2b So-yto 5a Su-yto 6 Su-to
Su-ybe 1 Su-be 2b So-be
Su-ybi 1 Su-bi 2b So-bi 5a Su-bi
Su-Wa 2b So-Wa 14b So-a

Stem: /di/ ‘give’

di-i 2b de-i 5a di-i 13 di-y
di-uk 2b de-uk 5a di-uk 13 di-wk 14a di-k
di-e 2b de-e 9 dE-e 13 dE-Y
di-en 2b de-en 9 dE-en 13 dE-Yn 14a dE-n
di-o 2b de-o 9 dE-o 10 da-o 13 da-W
di-Wa *in standard Bangla*: 2b de-Wa
in nonstandard Bangla: 2b de-Wa 9 dE-Wa
di-yo 2b de-yo 5a di-yo 14b di-o
di-yto 2b de-yto 5a di-yto 6 di-to
di-ybe 1 di-be 2b de-be
di-ybi 1 di-bi 2b de-bi 5a di-bi

Stem: /bOh/ ‘flow, carry’

bOh-i 5a boh-i 12 bo-i 13 bo-y
bOh-uk 5a boh-uk 12 bo-uk 13 bo-wk 14a bo-k
bOh-yo 2a bOyh-o 5a boyh-o 12 boy-o
bOh-yto 2a bOyh-to 5a boyh-to 12 boy-to
bOh-Wa 12 bO-Wa

Stem: /duh/ ‘milk’

duh-i 2b doh-i 5a duh-i 12 du-i 13 du-y
duh-uk 2b doh-uk 5a duh-uk 12 du-uk 13 du-wk 14a du-k
duh-yo 2ab doyh-o 5a duyh-o 12 duy-o
duh-yto 2ab doyh-to 5a duyh-to 12 duy-to
duh-Wa 2b doh-Wa 12 do-Wa 14b do-a

Stem: /gah/ ‘sing’

gah-i 12 ga-i 13 ga-y
gah-uk 12 ga-uk 13 ga-wk 14a ga-k
gah-yo 2a gayh-o 3 gEyh-o 5 geyh-o 12 gey-o
gah-yto 2a gayh-to 12 gay-to
gah-Wa 12 ga-Wa

Stem: /pat/ ‘spread’

pat-i 0 pat-i
pat-uk 0 pat-uk
pat-yo 2a payt-o 3 pEyt-o 5a peyt-o 6 pet-o
pat-yto 2a payt-to 6 pat-to
pat-Wa 8 pat-a

Stem: /pa/ ‘get’

pa-i 13 pa-y
pa-uk 13 pa-wk 14a pa-k
pa-yo 3 pE-yo 5a pe-yo
pa-yto 3 pE-yto 5a pe-yto 6 pe-to
pa-Wa 0 pa-Wa

Stem: /cap-Wa/ ‘impose’

cap-Wa-i 8 cap-a-i 13 cap-a-y
cap-Wa-uk 8 cap-a-uk 13 cap-a-k 14a cap-a-k
cap-Wa-yo 3 cap-WE-yo 4 cap-We-yo 5a cap-Wi-yo 8 cap-i-yo 14b cap-i-o
cap-Wa-yto 6 cap-Wa-to 8 cap-a-to

Stem: /tEWRWa/ ‘warp’

tEWRWa-uk 8 tEWRa-uk 13 tEWRa-wk 14a tEWRa-k
tEWRWa-yo 3 tEWRWE-yo 4 tEWRWe-yo 5a tEWRWi-yo 5a tEWRwi-yo 5a tEwRi-yo 5a tewRwi-yo 8 tewRi-yo 14b tewRi-o
tEWRWa-ye 3 tEWRWE-ye 4 tEWRWe-ye 5a tEWRWi-ye 5a tEWRwi-ye 5a tEwRwi-ye 5a tewRwi-ye 7 tewRwy-ye 8 tewRy-ye 8 tewR-ye 8 tewR-e
tEWRWa-yto 6 tEWRWa-to 8 tEWRa-to

Stem: /bayrWa/ ‘go out, come out’

bayrWa-i 3 bEyrWa-i 5ab beyrWo-i 6 berWo-i 8 bero-i 13 bero-y
bayrWa-uk 3 bEyrWa-uk 5ab beyrWo-uk 6 berWo-uk 8 bero-uk 13 bero-wk 14a bero-k
bayrWa-yo 3 bEyrWE-yo 4 bEyrWe-yo 5a beyrWi-yo 5a beyrwi-yo 6 berwi-yo 8 beri-yo 14b beri-o
bayrWa-ye 3 bEyrWE-ye 4 bEyrWe-ye 5a beyrWi-ye 5a beyrwi-ye 6 berwi-ye 8 beri-ye 14b beri-e

bayrWa-yto 3 bEyrWa-yto 5ab beyrWo-yto 6 berWo-to 8 bero-to

Stem: /kha-Wa/ ‘feed’

kha-Wa-i 13 kha-Wa-y
kha-Wa-uk 13 kha-Wa-wk 14a kha-Wa-k
kha-Wa-yo 3 kha-WE-yo 4 kha-We-yo 5a kha-Wi-yo 5a kha-wi-yo 7 kha-wy-yo 14a kha-y-yo
kha-Wa-ye 3 kha-WE-ye 4 kha-We-ye 5a kha-Wi-ye 5a kha-wi-ye 7 kha-wy-ye 14a kha-y-ye
kha-Wa-yto 6 kha-Wa-to

Stem: /di-Wa/ ‘cause to give’ (nonstandard output: *dE-Wa-…*)

di-Wa-i 2b de-Wa-I 13 de-Wa-y
di-Wa-uk 2b de-Wa-uk 13 de-Wa-wk 14a de-Wa-k
di-Wa-yo 2b de-Wa-yo 3 de-WE-yo 4 de-We-yo 5a de-Wi-yo 5a de-wi-yo 5a di-wi-yo 7 di-wy-yo 14a di-y-yo
di-Wa-yto 2b de-Wa-yto 6 de-Wa-to

Stem: /gah-Wa/ ‘cause to sing’

gah-Wa-i 12 ga-Wa-i 13 ga-Wa-y
gah-Wah-uk 12 ga-Wa-uk 13 ga-Wa-wk 14a ga-Wa-k
gah-Wa-yo 3 gah-WE-yo 4 gah-We-yo 5a gah-Wi-yo 5a gah-wi-yo 7 gah-wy-yo 12 ga-wy-yo 14a ga-y-yo
gah-Wa-yto 6 gah-Wa-to 12 ga-Wa-to

Stem: /bOh-Wa/ ‘cause to carry’

bOh-Wa-ye 3 bOh-WE-ye 4 bOh-We-ye 5a bOh-Wi-ye 5a bOh-wi-ye 5a boh-wi-ye 7 boh-wy-ye 12 bo-wy-ye 14a bo-y-ye

2.3.2 The aspected forms

An aspected finite verb form consists of stem plus nonfinite ending (conjunctive or infinitive) plus auxiliary plus tense plus finite ending. The auxiliary must be /ch/ or /thak/. If /ch/, and if the stem preceding it is in the infinitive, then this infinitive ending has the shape /yc/, and Stem-yc-ch-(Tense)-Ending is a single word. Otherwise, there is a word boundary before the auxiliary (a phonological word boundary, shown as # in the chart below; conventional orthography does not show this boundary, but writes Verb-ye-ch-(Tense)-Ending as a single word). Here is a paradigm for the stem /kOr/.

	3P NEUTR	3-2P HON	2P NEUTR	2P INTIM	1P
IND PRES PERF	kore#che	kore#chen	kore#cho	kore#chiS	kore#chi
IMP PRES PERF	kore :thakuk	:thakun	:thako	:thakiS	:thaki
IND PRES PROG	korche	korchen	korcho	korchiS	korchi

IMP PRES PROG	korte :thakuk	:thakun	:thako	:thakiS	:thaki
PAST PERF	kore#chilo	kore#chilen	kore#chile	kore#chili	kore#chilam
PAST PROG	korchilo	korchilen	korchile	korchili	korchilam
FUT PERF	kore :thakbe	:thakben	:thakbe	:thakbi	:thakbo
FUT PROG	korte :thakbe	:thakben	:thakbe	:thakbi	:thakbo
COND PERF	kore :thakto	:thakten	:thakte	:thakti	:thaktam
COND PROG	korte :thakto	:thakten	:thakte	:thakti	:thaktam
NEUTR PERF	kore :thake	:thaken	:thako	:thakis	:thaki
NEUTR PROG	korte :thake	:thaken	:thako	:thakiS	:thaki

An aspected non-finite verb form consists of stem plus non-finite ending (conjunctive or infinitive) plus auxiliary /thak/ plus non-finite ending. In principle, /kOr/ could yield the Perfect forms *kore theke, kore thakte, kore thakle* and the Progressive forms *korte theke, korte thakte, korte thakle*. But the double Conjunctive *kore theke* (underlying form /kOr-ye thak-ye/) is completely ungrammatical, and the double Infinitive *korte thakte* (underlying form /kOr-yte thak-yte/) is nearly as bad. This leaves four occurring forms: Perfect *kore thakte* 'to have done' and *kore thakle* 'if...has done', and Progressive *korte theke* 'having been doing' and *korte thakle* 'if... is doing'.

An aspected nominal verb form (an aspected *participle-gerund*, as we shall call it in the more careful sections of this description, later) consists of stem plus non-finite ending (conjunctive or infinitive) plus auxiliary /thak/ plus ending /Wa/. The stem /kOr/ yields the Perfect form *kore thaka* 'having done' and the Progressive form *korte thaka* 'to be doing'.

Here are some phonological derivations.

Stem: /kOr/ 'do'

kOr-yc-ch-e 2a kOyr-c-ch-e 5a koyr-c-ch-e 6 kor-c-ch-e 11 kor-ch-e
kOr-ye#ch-e 2a kOyr-e#ch-e 5a koyr-e#ch-e 6 kor-e#ch-e

Stem: /tul/ 'lift'

tul-yc-ch-e 2ab toyl-c-ch-e 5a tuyl-c-ch-e 6 tul-c-ch-e 11 tul-ch-e

Stem: /hO/ 'be, become'

hO-yc-ch-e 5a ho-yc-ch-e 6 ho-c-ch-e
hO-ye#ch-e 5a ho-ye#ch-e

Stem: /Su/ 'lie down'

Su-yc-ch-e 2b So-yc-ch-e 5a Su-yc-ch-e 6 Su-c-ch-e
Su-ye#ch-e 2b So-ye#ch-e 5a Su-ye#ch-e

Stem: /bOh/ 'flow, carry'

bOh-yc-ch-e 2a bOyh-c-ch-e 5a boyh-c-ch-e 11 boyh-ch-e 12 boy-ch-e
bOh-ye#ch-e 2a bOyh-e#ch-e 5a boyh-e#ch-e 12 boy-e#ch-e

Stem: /duh/ 'milk'

duh-yc-ch-e 2ab doyh-c-ch-e 5a duyh-c-ch-e 11 duyh-ch-e 12 duy-ch-e
duh-ye#ch-e 2ab doyh-e#ch-e 5a duyh-e#ch-e 12 duy-e#ch-e

Stem: /gah/ ‘sing’
gah-yc-ch-e 2a gayh-c-ch-e 11 gayh-ch-e 12 gay-ch-e
gah-ye#ch-e 2a gayh-e#ch-e 3 gEyh-e#ch-e 5a geyh-e#ch-e 12 gey-e#ch-e

Stem: /pat/ ‘spread’
pat-yc-ch-e 2a payt-c-ch-e 6 pat-c-ch-e 11 pat-ch-e
pat-ye#ch-e 2a payt-e#ch-e 3 pEyt-e#ch-e 5 peyt-e#ch-e 6 pet-e#ch-e

Stem: /pa/ ‘get’
pa-yc-ch-e 6 pa-c-ch-e
pa-ye#ch-e 3 pE-ye#ch-e 5a pe-ye#ch-e

Stem: /bayrWa/ ‘go out, come out’
bayrWa-yc-ch-e 3 bEyrWa-yc-ch-e 5ab beyrWo-yc-ch-e 6 berWo-c-ch-e 8 bero-c-ch-e
bayrWa-ye#ch-e 3 bEyrWE-ye#ch-e 4 bEyrWe-ye#ch-e 5a beyrWi-ye#ch-e
5a beyrwi-ye#ch-e 6 berwi-ye#ch-e 8 beri-ye#ch-e 14b beri-e#ch-e

Stem: /kha-Wa/ ‘feed’
kha-Wa-yc-ch-e 6 kha-Wa-c-ch-e
kha-Wa-ye#ch-e 3 kha-WE-ye#ch-e 4 kha-We-ye#ch-e 5a kha-Wi-ye#ch-e
5a kha-wi-ye#ch-e 7 kha-wy-ye#ch-e 14a kha-y-ye#ch-e

2.3.3 The mannered forms

A mannered verb form consists of stem plus ending *ye* plus # plus one of over twenty special stems plus any sequence which added to a verb stem yields a simple form or an aspected form. A mannered verb is usually called a Compound Verb, the second stem being called a Vector or a Vector Verb Stem. Dasgupta 1977 has proposed that the first stem of a compound verb be called its Pole, and that the term Pole also refer to the (sole) stem in a non-compound verb (simple or aspected). All verb stems that function as vectors can also function (in other sentences, of course) as poles.

Nothing need be said morphologically about compound verb forms. The first word in such a form (in its canonical word order) is a simple non-finite form; the residue of the compound verb is either simple or aspected, and must in addition choose between a finite ending, an adverbial non-finite ending, and a nominal (participle-gerund) ending; brief descriptions and examples of all these inflectional forms of V are given above.

2.4 Lexical categories

On the basis of the above one may say, as traditional grammarians did, that on morphological grounds one should set up for Bangla at least two major word classes: Nouns, declined, and Verbs, conjugated. We will not offer any stronger arguments for considering Noun and Verb as Lexical Categories in the grammar of Bangla. Particular stems in the lexicon will be marked as

N or V; e.g. /kagoj/ ‘paper’ is an N and /tEWRWa/ ‘warp’ a V.

It will be assumed, without specific discussion, that the notions N and V can be decomposed into features of some sort. As mentioned in chapter 1, this inquiry assumes the general theoretical framework of lexicalist generative grammar as initiated in Chomsky 1970.

Beyond these rudimentary points, difficult questions arise, especially concerning (i) the constituent structure of phrases which contain N words and those which contain V words, (ii) the relation between morphology and syntax, and (iii) the internal structure of noun phrases like *jeTa* ‘which one’ touched upon in section 2.2.6.

Until such issues are formulated and in part resolved, one must work with what may seem obvious principles but are in fact unsupported strategic assumptions. These assumptions, some of which are stated at the beginning of chapter 3, determine what sorts of Phrase Structure rules this grammar uses and how it conceives of lexical entries.

Chapter 3

OVERALL BANGLA GRAMMAR: N AND V IN PHRASE STRUCTURE

3.0 Strategy and tactics

3.0.1 Strategy

I continue the task, begun in chapter 2, of sketching an overall grammar for the language. Chapter 2 dealt with some fragments of the basic phonology and morphology. From chapter 4 on I will be concerned with syntactic and semantic problems. The present chapter examines the border zone between linguistics inside the word (phonology and morphology) and linguistics between words (syntax and semantics). The reason for this examination is that, as chapter 1 says, the internal structure of relative clauses and constituent questions is one of the things under investigation here, and one aspect of this structure is the morphosyntactic composition of relative and interrogative NPs. From this point of view the most important result obtained in chapter 3 is that relative 'pronouns' like *je* 'who', and interrogative 'pronouns' like *ke* 'who?' consist of an overt determiner followed by a zero noun: *je Ø*, *ke Ø*. It is the DET node that carries the relative or interrogative element (the "open" element), just as in fuller NPs like *je raja* 'which king' or *kon raja* 'which king?'. The discussion of verbs later in the chapter is a necessary prelude to the analysis of verbal, especially infinitival, structures in the following chapters. Infinitivals, in turn, are of interest because a study of relative and interrogative constructions (open constructions) needs to address the question as to why infinitival structures, which are in some sense propositional, do not in Bangla act in a relative or interrogative role.

3.0.2 Tactics

I assume here P.H. Matthews' word-and-paradigm approach to inflectional morphology linked to other realms of grammar in the manner specified by Aronoff (1976). The main advantage of this approach is that it confines the cases of non-phonologically determined allomorphy to morphology, leaving syntactic processes proper independent of facts about allomorphy.

In my attempts to implement the Matthews-Aronoff approach to problems of inflectional morphology I have, as far as possible, made specific choices of format for presenting lexical entries etc. only to the extent that the data considered here force me into such choices. This seems to me appropriate in a first overall generative grammar of a language. The present volume is best regarded as a collection of small-scale studies of selected aspects of Bangla grammar – a collection which, to the degree required by the main problem of the dissertation, does cohere and yield systematic predictions, but which neither answers all the questions one might reasonably ask about the precise form and content of the grammatical theory presupposed by this grammar nor fills in all the gaps between, say, the morphology and the syntax presented. This particular lacuna could be removed if a well-built general theory of the morphology-syntax interface were available.

This chapter presents a first approximation to a theory of NP and VP structure. The terrains considered are those morphologically charted in chapter 2. The analysis of VP structure runs on into chapters 4 and 5.

3.1 Nouns and Determiners

3.1.1 Nouns and pronouns

For ease of exposition, chapter 2 made the traditional assumption that all pronouns are Nouns. Section 2.2.6 showed that *Ta* in *boy Ta* 'the book' is a DENOM and that the form derives from the N NUML DENOM sequence *boy Ek Ta* "book one item" by the application of a rule deleting *Ek*. This analysis of N + *Ta* subsumes pronominal forms like *konTa* 'which one?' if all pronouns are Nouns.

I will now show that pronouns that can precede *Ta* are not nouns but DET N sequences where N is a silent or zero noun. This demonstration keeps two promises made in chapter 2: the promise to show that some (including relative and interrogative) pronouns are not nouns and the promise to revise slightly chapter 2's account of forms ending in *Ta*. The nature of the revision is as follows. Chapter 2 suggested, without stating it, that pronominal forms like *konTa* 'which one?' are N + *Ta* sequences derived by *Ek*-Deletion from N *Ek Ta*. It will now be shown that PRONOUN *Ta* sequences in fact have the surface structure DET N *Ta* which *Ek*-Deletion derives from Det N *Ek Ta*.

I begin by recapitulating some results obtained in chapter 2. Then I present two options for the categorial status of the relative pronoun *je*: (i) that it is an N and (ii) that it is a DET. A confrontation with the relevant data shows that (ii) is the better option. The choice of *je* from the group *e o Se kono kon je* is arbitrary in that the demonstration would have gone through for the others as well; but deliberate in that the reason why the categorial status of these elements interests us is that we want to know the detailed structure of relative clauses and constituent questions, and such clauses contain respectively relative and interrogative pronominal elements.

3.1.2 Recapitulation of some NP structure

Chapter 2 posited the phrase structure rules repeated as (1) or (2) and proposed a movement rule, (3), as one of two equally good ways of handling a certain class of NP data. The output of (3) was to undergo (4), a deletion rule whose obligatoriness explained ill-formedness of, say, **lok Ek Ta* "person one item" and **lok kOtok gulo* "person some GULO".

(1) NP → (NUM′) N
(2) NUM′ → NUML DENOM
(3) NUM′ Postposing, optional
S.D.: NUM′ – N, where NUM′ contains *Ta* or *khana* or *gulo*
S.C.: 1 2 ➔ 0, 2+1
(4) NUML-Deletion, obligatory
S.D.: N – NUML – DENOM, with various conditions on the terms
S.C.: 1 2 3 ➔ 1, 0, 3

The conditions on the S.D. of (4) would be roughly as follows. Term 3 must be *Ta* or *khana* or

gulo. If term 3 is *gulo,* term 2 must be *kOtok.* If term 3 is *Ta* or *khana,* term 2 must be *Ek.* It was tentatively assumed that *ra, der, ke, er* were declensional suffixes and that *khana, Ta, jon, peYala, gOj, koTi, gulo* etc. were Denominators.

3.1.3 One difficulty

These proposals are observationally inadequate if conjoined with the assumption that pronouns are Nouns. For then *jeTa* 'which one, the one which' must have the following derivation.

(5) a. NP
b. NUM′ N
c. NUML DENOM N
d. Ek Ta je
e. je Ek Ta
f. je Ta

The problem is that, since rule (3), which derives line *e* from line *d*, is optional, *d* should be a well-formed string with the surface structure $[_{NP}\ [_{NUM'}\ [_{NUML}$ Ek$][_{DENOM}$ Ta$]][_{N}$ je$]]$. But this surface structure is ill-formed. In general $[_{NP}$ NUM′ Y$]$ is ill-formed whenever Y is a pronoun. Lexical entries for pronouns should have (or should get, by some general rule) a subcategorization feature [–NUM′__] making it illegitimate to derive (5-d) from (5-c). Some derivation other than (5) is called for.

I do not wish to be too hasty. Before exploring other alternatives (in section 3.1.4), let me dwell on (5) and ways in which one might try to 'save' it. Two ways to do this come to mind.

(A) One might keep (5) intact; not assign the feature [–NUM′__] to pronouns, and devise a surface structure filter that stars the sequence NUM′ Pronoun.

(B) One might revise (5) to the not very different (6), which assumes a model where lexical insertion rules apply after all movement rules. This idea, known as 'surface lexicalization', has been proposed off and on in lexicalist linguistics. If (6) is taken to be the derivation of *jeTa,* then one may assign pronouns the subcategorization feature [–NUM′__] if one wishes.

(6) a. NP
b. NUM′ N
c. NUML DENOM N
d. N NUML DENOM
e. je Ek Ta
f. je Ta

In what may be the only comprehensive study so far of problems of lexical insertion in a transformational grammar, Seegmiller (1974) has argued rather cogently against surface lexicalization. Chapter 10 of this volume adds to his arguments. So, we need not take (B) seriously.

(A) is less easy to dismiss out of hand. But several considerations count against it. There seem to be none in its favour.

First, the proposed filter does not meet the formal conditions on surface filters suggested in Chomsky & Lasnik (1977). It does not involve COMP or clause structure. It mentions an entity ('pronoun') which is neither a specific formative nor a categorial notion in any theory of categories. Chomsky & Lasnik's are the only suggestions on record as to what conditions filters should meet.

Second, the proposed filter is redundant, for there are other grounds (as we shall see) on which one can rule out surface structures of the form NUM′ Pronoun. Chomsky (1978) suggests that, if some structure is ungrammatical, linguists should try to attribute its ungrammaticality to one reason rather than several, wherever possible. By this principle we should dispense with (A), given the availability of independent means to achieve its effect.

Third, the filter *NUM′ Pronoun is loosely stated. What one actually needs ruled out is $[_{NP}$ NUM′ (AP^n) Pronoun], where AP^n is a sequence of one or more Adjective Phrases. With this added complexity the facts can still be accommodated in a subcategorization feature analysis, since Chomsky (1965) explicitly provided for this property of the subclass of subcategorization features that he called selectional features – he provided for their ability to take effect 'over a variable'. But filters never apply over a variable. If they are allowed to apply over a variable, the conceptual difference between Bresnan's non-filter analyses and Chomsky's filter analyses disappears. Thus, for reasons of principle, we must use a lexical analysis of the facts at issue, not one in terms of a filter.

Having shown that such attempts to 'save' (5) are ill-advised, I will now build and assess alternative hypotheses.

3.1.4 A proposed solution schema

Since (5-d), the putative source for (5-e), apparently should not be base-generated, let us go about base-generating (5-e). At this point, it is advisable to stop feeling sure that *je* is a noun. For if we assume that *je* is a noun, then in order to base-generate (5-e) we must posit a rule like NP → N NUM′. This is a possibility. But nothing else we know about N so far compels us to accept the idea. Let us, then, withdraw our assumption that *je* is a noun. Using U for Unknown, let us assign to (5-e) the structure $[_{NP}\,[_{U}$ je$][_{NUM'}\,[_{NUML}$ Ek$][_{DENOM}$ Ta$]]]$. To get a broader idea of the behavior of U, consider (7) and (8), where the use of bold type at (7vii) indicates intonational prominence.

(7)
i. je lok 'which person'
ii. je lok Ta 'which person'
iii. je du To lok 'which two people'
iv. je lok du To 'which two people'
v. je tin Te lok 'which three people'
vi. je lok tin Te 'which three people'
vii. je Ek Ta lok 'which one person'
viii. *je lok Ek Ta
ix. *je Ta lok

(8) i. lok ‘(a) person’
ii. lok Ta ‘the person’
iii. du To lok ‘two people’
iv. lok du To ‘the two people’
v. tin Te lok ‘three people’
vi. lok tin Te ‘the three people’
vii. Ek Ta lok ‘a person, one person’
viii. *lok Ek Ta
ix. *Ta lok

If we ignore the phonological and semantic emphasis in (7-vii) not matched by (8-vii), we can make a general assertion: *je* + (8) = (7). But (8) displays the full range of NP structures examined so far: the range generated by rules (1) and (2). So, in order to accommodate (7), we modify (1) to (9), and we keep (2).

(9) NP → (U) (NUM′) N
(2) NUM′ →NUML DENOM

Phrase structure rule (9) has an optional U node, U for Unknown. If we select this node, the result – also shaped by rule (2) and the transformations stated earlier – will be one of the forms in (7). The corresponding form in (8) will result if the U node is not selected when applying (9).

Without yet asking what U is, let us look at the difference between two relevant derivations – those of (7-vi) and (8-vi).

(10) a. NP
b. U NUM′ N
c. U NUML DENOM N
d. je tin Ta lok
e. je lok tin Ta

(11) a. NP
b. NUM′ N
c. NUML DENOM N
d. tin Ta lok
e. lok tin Ta

In this sort of derivation, at least, U can be N as was initially assumed. If U is indeed N, the derivation goes through without a hitch. But U could also not be N and the derivation would still go through. Proposal (9), compatible with either possibility, is a schema for solving the problem of (5), not itself a solution. Let us now formulate two solutions for (5) which differ in the values they assign to U in (9).

3.1.5 Two realizations of the schema

U must either be a Head element in the NP (an option which we will represent as N) or a

Dependent element (we will represent this as DET). Joanne Grumet has pointed out to me that a variant of the second option (a variant which she would represent as ADJ) neatly fits the data of the Indic language Romany. However, it seems necessary to keep DET and ADJ distinct in general, so that one can make statements like 'DET and ADJ follow N in Bahasa Indonesia and precede N in English, but in French DET precedes N while ADJ follows N'. On these rather tenuous grounds I have decided to use the term DET and not ADJ for the value of U in this study, despite the traditional term 'Demonstrative Adjectives' for some of the items I am calling 'Determiners'. There is a question of theoretical principle here which will have to be untangled and solved in future work. The question has to do with how Specifiers and Modifiers differ. Presumably a Specifier, unlike a Modifier, closes its phrase.

One might wish to state the hypotheses as simply (12) and (13).

(12) NP → (N) (NUM′) N
(13) NP → (DET) (NUM′) N

However, the difference is actually more complicated. Consider $[_{NP}\,[_{U}$ je$][_{NUM'}$ Ta$]]$, where there is no second N. Now, if U is N, there need not be a second N here. The first N can function as the head of the NP. So (12) may be revised to (14). But if U is DET, the NP must have an N to U's right in order to satisfy the generalization that NPs are endocentric. The rule must then be stated as (15) – i.e. as (13) plus the possibility that N may dominate a terminal element of phonologically unrealized and otherwise not yet defined character.

(14) The U-is-N Hypothesis: NP → (N) (NUM′) (N)
(15) The U-is-DET Hypothesis: NP → (DET) (NUM′) N, where N may be 'empty' in some sense

Of course, a proponent of U-is-N could just as well go in for the (12) plus 'empty' N option not considered here. The point is not that a proponent of U-is-N cannot do this. The point is that, if proponents of both hypotheses agree that an NP should have at least one N in it, then, following the path of least resistance, a believer in U-is-N can posit (14) and leave it up to some convention to ensure that at least one of the N nodes gets selected, whereas a proponent of U-is-DET is led under the same circumstances to choose (15) rather than, say, NP → (DET) (NUM′) (N). It is not a matter of what logically must be done, but of what, given the shared method, would normally be done.

3.1.6 Comparing the alternatives: neutral examples

Over a wide range of data, (14) and (15) make the same predictions as far as empirical testing can judge. In some of these cases, (14) and (15) predict different numbers of phrase-markers, but such differences are untestable by present-day methods.

I begin by providing derivations of examples for which (14) and (15) make identical predictions. Then I turn to cases of logically different but objectively indistinguishable predictions, and finally, in section 3.1.7, to crucial facts, which select (15) over (14).

For the examples cited in (16), the two hypotheses make identical predictions. More explicitly, for each of these cases both hypotheses assign the same values for *n* and *s* in the

statement-schemata ‘this example has *n* structures and *s* grammaticality’, where *n* is an integer and *s* is plus if the string has at least one well-formed structure and is minus otherwise. Some of the derivations for (16) are given in (17) (I omit the derivations for (v) and (vi) as they are exactly parallel to (iii) and (iv)). In all the derivations given, the first phrase-marker results from applying lexical insertion to the output of the categorial rule mentioned before the phrase-marker.

(16) i. je lok ‘which person [= the person who]’
ii. je lok Ta ‘which person’
iii. je du To lok ‘which two people’
iv. je lok du To ‘which two people’
v. je tin Te lok ‘which three people’
vi. je lok tin Te ‘which three people’
vii. *je lok Ek Ta
viii. je ‘who’
ix. je Ta ‘which one’

(17) i. 1. by (14): $[_{NP}[_{N}$ je$][_{N}$ lok$]]$
2. by (15): $[_{NP}[_{DET}$ je$][_{N}$ lok$]]$
ii. 1. by (14): a. $[_{NP}[_{N}$ je$][_{NUM'}[_{NUML}$ Ek$][_{DENOM}$ Ta$]][_{N}$ lok$]]$
by (3): b. $[_{NP}[_{N}$ je$][_{N}$ lok$][_{NUM'}[_{NUML}$ Ek$][_{DENOM}$ Ta$]]]$
by (4): c. $[_{NP}[_{N}$ je$][_{N}$ lok$][_{NUM'}[_{DENOM}$ Ta$]]]$
2. by (15): a. $[_{NP}[_{DET}$ je$][_{NUM'}[_{NUML}$ Ek$][_{DENOM}$ Ta$]][_{N}$ lok$]]$
by (3): b. $[_{NP}[_{DET}$ je$][_{N}$ lok$][_{NUM'}[_{NUML}$ Ek$][_{DENOM}$ Ta$]]]$
by (4): c. $[_{NP}[_{DET}$ je$][_{N}$ lok$][_{NUM'}[_{DENOM}$ Ta$]]]$
iii. 1. by (14): $[_{NP}[_{N}$ je$][_{NUM'}[_{NUML}$ du$][_{DENOM}$ Ta$]][_{N}$ lok$]]$
2. by (15): $[_{NP}[_{DET}$ je$][_{NUM'}[_{NUML}$ du$][_{DENOM}$ Ta$]][_{N}$ lok
iv. 1. by (14): a. $[_{NP}[_{N}$ je$][_{NUM'}[_{NUML}$ du$][_{DENOM}$ Ta$]][_{N}$ lok$]]$
by (3): b. $[_{NP}[_{N}$ je$][_{N}$ lok$][_{NUM'}[_{NUML}$ du$][_{DENOM}$ Ta$]]]$
2. by (15): a. $[_{NP}[_{DET}$ je$][_{NUM'}[_{NUML}$ du$][_{DENOM}$ Ta$]][_{N}$ lok$]]$
vii. 1. by (14): a. $[_{NP}[_{N}$ je$][_{NUM'}[_{NUML}$ Ek$][_{DENOM}$ Ta$]][_{N}$ lok$]]$
by (3): b. $[_{NP}[_{N}$ je$][_{N}$ lok$][_{NUM'}[_{NUML}$ Ek$][_{DENOM}$ Ta$]]]$
by convention: c. assign *: obligatory rule (4) has failed to apply
2. by (15): a. $[_{NP}[_{DET}$ je$][_{NUM'}[_{NUML}$ Ek$][_{DENOM}$ Ta$]][_{N}$ lok$]]$
by (3): b. $[_{NP}[_{DET}$ je$][_{N}$ lok$][_{NUM'}[_{NUML}$ Ek$][_{DENOM}$ Ta$]]]$
by convention: c. assign *: (4) hasn’t applied
viii. 1. by (14): $[_{NP}[_{N}$ je$]]$
2. by (15): $[_{NP}[_{DET}$ je$][_{N}\ ]]$
ix. 1. by (14): a. $[_{NP}[_{N}$ je$][_{NUM'}[_{NUML}$ Ek$][_{DENOM}$ Ta$]]]$
by (4): b. $[_{NP}[_{N}$ je$][_{NUM'}[_{DENOM}$ Ta$]]]$
2. by (15): a. $[_{NP}[_{DET}$ je$][_{NUM'}[_{NUML}$ Ek$][_{DENOM}$ Ta$]][_{N}\ ]]$
by (3): b. $[_{NP}[_{DET}$ je$][_{N}\ ][_{NUM'}[_{NUML}$ Ek$][_{DENOM}$ Ta$]]]$
by (4): c. $[_{NP}[_{DET}$ je$][_{N}\ ][_{NUM'}[_{DENOM}$ Ta$]]]$

In (17-vii), ungrammaticality is assigned because (4), an obligatory rule, has failed to apply.

For the examples cited in (18), the two hypotheses assign different values to *n* (but not to *s*) but the differences cannot be checked empirically, since only *s* is observable. (19) gives the derivation for (18-i) and the 'non-derivations' for (18-ii, iii).

(18)i. je du To 'those two (items)'
ii.*Ek Ta je
iii.*Ta je

(19) i. 1. by (14): $[_{NP}\,[_{N}$ je$][_{NUM'}\,[_{NUML}$ du$][_{\ DENOM}$ Ta$]]]$
2. by (15): $[_{NP}\,[_{DET}$ je$][_{NUM'}\,[_{NUML}$ du$][_{\ DENOM}$ Ta$]][_{N}\ ]]$
3. by (15): a. $[_{NP}\,[_{DET}$ je$][_{NUM'}\,[_{NUML}$ du$][_{\ DENOM}$ Ta$]][_{N}\ ]]$
by (3): b. $[_{NP}\,[_{DET}$ je$][_{N}\ ][_{NUM'}\,[_{NUML}$ du$][_{\ DENOM}$ Ta$]]]$

ii. 1. by (14): $[_{NP}\,[_{NUM'}\,[_{NUML}$ Ek$][_{\ DENOM}$ Ta$]][_{N}$ je$]]$; but in this environment, an instance of NUM′ ___, *je* would fail to be lexically inserted.
2. as if by (15): $[_{NP}\,[_{NUM'}\,[_{NUML}$ Ek$][_{\ DENOM}$ Ta$]][_{DET}$ je$][_{N}\ ]]$ ungenerable because no sequence of rules including (15) would generate NUM′ to the left of DET
3. as if by (15): $[_{NP}\,[_{NUM'}\,[_{NUML}$ Ek$][_{\ DENOM}$ Ta$]][_{N}\ ][_{DET}$ je$]]$ ungenerable for the same reason as (ii-2)

iii. 1. as if by (14): $[_{NP}\,[_{NUM'}[_{DENOM}$ Ta$]][_{N}$ je$]]$ ungenerable because no sequence of rules including (14) would generate $[_{NP}\,[_{NUM'}$ DENOM]N]
2. as if by (15): $[_{NP}\ [_{NUM'}[_{DENOM}$ Ta$]][_{DET}$ je$][_{N}\ \ ]]$ ungenerable because no sequence of rules including (15) would generate NUM′ to the left of DET
3. as if by (15): $[_{NP}\ [_{NUM'}[_{DENOM}$ Ta$]][_{N}\ \ ][_{DET}$ je$]]$ ungenerable for the same reason as (iii-2)

In each of the above cases, hypothesis (15) generates (or 'would generate', in the case of the ill-formed strings) two structures where hypothesis (14) generates or 'would generate' only one. But the two structures always go the same way – both grammatical or both ungrammatical. Thus, in each case, (15) ends up not making a checkably different prediction from ((14).

3.1.7 Comparing the alternatives: crucial examples

I turn now to two sets of crucial examples which support hypothesis (15) against (14). In the first set of data, (20), for which (21) gives the predictions in detail, it is necessary to assume that the contrastive stress on *Ek* in *je Ek Ta (lok)* does not affect its syntactic status. Some may find this assumption unpalatable. For such readers the second set of data, (22), and the predictions about them in (23), will still clinch the issue in favour of (15).

(20) i. je Ek Ta lok 'which one person'
ii. je Ek Ta 'which one (item)'

(21) i. 1. by (14): a. $[_{NP}\,[_{N}$ je$][_{NUM'}\,[_{NUML}$ Ek$][_{\ DENOM}$ Ta$]][_{N}$ lok$]]$
by convention: b. assign *: (4) hasn't applied
2. by (15): $[_{NP}\,[_{DET}$ je$][_{NUM'}\,[_{NUML}$ Ek$][_{\ DENOM}$ Ta$]][_{N}$ lok$]]$

ii. 1. by (14): a. [NP [N je][NUM′ [NUML Ek][DENOM Ta]]]
by convention: b. assign *: (4) hasn't applied
2. by (15): [NP [DET je][NUM′ [NUML Ek][DENOM Ta]][N]]

(22) i. je du jon 'which two (people), = the two people who': the crucial case
ii. je du jon lok 'which two people'
iii. *je lok du jon

(23) i. 1. by (14): a. [NP [N je][NUM′ [NUML du][DENOM jon]]]
by convention: b. assign *; no semantic rule interprets N NUML DENOM
if DENOM is *jon* (cf. (22-iii))
2. by (15): [NP [DET je][NUM′ [NUML du][DENOM jon]][N]] interpretation goes through (cf. (22-ii))
3. by (15): a. [NP [DET je][NUM′ [NUML du][DENOM jon]][N]]
by (3): b. [NP [DET je][N][NUM′ [NUML du][DENOM jon]]]
by convention: c. assign *; no semantic rule interprets N NUML DENOM
if DENOM is *jon* (cf. (22-iii))

Hypothesis (14) incorrectly predicts that *je du jon* should be ill-formed. (15) correctly predicts that *je du jon* should be well-formed (with one structure). Thus, the fact that *je du jon* does occur selects (15) over (14).

3.1.8 Refining the proposal: phonologically empty nouns

It is time to define 'empty N' now – to specify the exact content of what was represented as [N] in the derivations. Two definitions come to mind. The choice between them proves to be an empirical one.

ENV, the Empty Node View: Lexical insertion rules are optional. Where the insertion process exercises its option of not applying and leaves a node dominating no terminal material, that slot is conventionally filled by the so-called concatenational element often written *e* for empty. The empty noun of (15) is such an *e* dominated by N.

ZIV, the Zero Item View: One or more lexical items exist which, though phonologically blank (hence representable as Ø), do have syntactic and semantic features. The empty noun of (15) is such a zero noun.

I will discuss the substance of ENV and ZIV first and turn to their pedigree later.

Empirical arguments can be given in support of ZIV on the basis of crucial data not presented here. See the Appendix to this chapter for details. The relevant data and argumentation are complicated. The following theoretical argument in favour of ZIV is less involved than the empirical arguments in the appendix and relies on simple and widely accepted premises. For some readers, this conceptual reasoning will suffice, and the exposition

can move on. Readers who wish to see the empirical arguments can consult the appendix before resuming their reading of this section.

Emonds (1970, 1976) prohibited surface structures containing empty nodes. This prohibition expressed the insight, which one might call Syntactic Verisimilitude, that every substructure in the correct surface structure/s for a well-formed sentence will serve some grammatical purpose. In Emonds' model, empty nodes in surface structure do not constitute input to any rule, and thus play no grammatical role. So in his model Verisimilitude is expressed most naturally by prohibiting surface structures that contain empty nodes. But in subsequent models of linguistic structure, empty nodes present in surface structure can undergo semantic interpretation (Wasow 1972, Jackendoff 1977, Chomsky & Lasnik 1977, Bresnan 1978). They are either placed under the anaphoric control of some non-empty node in the same sentence or assigned arbitrary reference. The appropriate way to express Verisimilitude insight in such models would be to formulate a principle requiring that all empty nodes undergo semantic interpretation which will assign reference either arbitrarily, or under the control of some other referring expression in the sentence. Thus, surface structure empty nodes in theories which allow them must carry heteronomous rather than autonomous reference.

It follows that a surface empty node must be the sort of node that can refer. The practice of grammarians who use empty nodes requires only that maximal phrase projections (NP, AP, etc.) refer. It is reasonable to postulate that only maximal phrase projections are referential nodes and that therefore only maximal phrase nodes like NP, AP, etc. may be empty at surface structure. If this principle, which merely formalizes current practice, is accepted, then empty Nouns, in the sense of an N dominating *e*, are not allowed at the level of surface structures.

Mark Baltin has pointed out to me that, in English, V is not a maximal phrase node, but that there is a vacant V site in the ( ) position in *Although I don't like steak, I do ( ) pizza*. This, however, would be a counterexample only if it were clear that one must postulate a surface V node for the ( ) position in such a string. Chomsky (1978) suggests that such strings result from deletion. For Chomsky & Lasnik (1977) and Chomsky (1978), deletion removes the node with its content. It is possible that the 'surface' structure which constitutes input to semantic interpretation contains the word *like (I do like pizza)* which is independently removed, node and all, by a deletion rule whose output is invisible to semantic rules. To my knowledge, there is no evidence that non-maximal nodes need to be able to be empty at the level of surface structure.

The above considerations and those in the appendix to this chapter strongly support ZIV against ENV. Readers who remain unconvinced may still concede the point in view of two things. First, there is no evidence to support ENV, nor, to my knowledge, any theoretical considerations which make ENV plausible. Second, there is evidence, given now, suggesting that not one but at least three distinct zero nouns are needed.

Semantic evidence: The zero noun which occurs in the position marked ____ in (24) and (25) consistently receives an inanimate count reading. Thus, *eTa* 'this one' can mean 'this door' in the appropriate context but never 'this lady'. By contrast, the zero noun that occurs in the position marked ____ in (26) is interpreted as animate: *je* 'who', *e* 'this person'. One can handle these facts quite naturally by positing at least two zero nouns, inanimate count C for (24) and (25) and animate A for (26):

(24) Det ___ Ta
(25) Det ___ gulo
(26) Det ___

Distributional evidence: Consider the portions of the pronoun chart which are repeated below as (27), (28) and (29):

(27)

GLOSS	NOM.SG	OBJ.SG	GEN.SG	LOC.SG
'that one'	SeTa	SeTake	SeTar	SeTate
'which one?'	konTa	konTake	konTar	konTate
'which one'	jeTa	jeTake	jeTar	jeTate

(28)

GLOSS	NOM.SG	OBJ.SG	GEN.SG	LOC.SG
'they$_5$'	Se	take	tar	--
'who?'	ke	kake	kar	--
'who$_1$'	je	jake	jar	--

(29)

'that$_2$ '	ta	take	tar	tate
'what?'	ki	kake	kiSer	kiSe
'what'	ja	jake	jar	jate

Morphic peculiarities distinguish inanimate count forms, animate forms, and inanimate mass forms from each other, as is readily apparent. The obvious way to account for this allomorphy is to postulate three distinct zero nouns: animate A, count inanimate C, and mass inanimate M. The determiners can then be said to take different shapes before different zero nouns. The great pre-structuralist linguist Chatterji (1968: 126-127) suggested regarding *ke* 'who?' and *kon* 'which?' as alternants of a single morpheme; the present proposal implements this suggestion of his in a systematic fashion.

Since a satisfactory morphology of Bangla animate and inanimate pronouns requires three zero nouns A, C, and M, it seems necessary to believe in zero nouns to begin with. This is a bit like arguing that the theory of Indo-European vowels requires at least three laryngeals A, E, and O, and that therefore one might as well concede the existence of laryngeals in the first place.

Let me close the discussion by giving some sources for ENV and ZIV. The choice between these ideas has been a potential issue since Chomsky (1965). Chomsky proposed a detailed theory of lexical insertion in two variants. In one variant, phrase structure rules introduced features under lexical category nodes, the task of the lexical rule being to insert every lexical item into a position whose features match the lexical item's features. In the other variant, phrase structure rules leave lexical category nodes dominating nothing; it is this variant that Seegmiller (1974) has described in detail and argued to be superior to the first variant on empirical grounds. But my formulations of ENV and ZIV borrow, not from Chomsky (1965), but respectively from Chomsky (1978) and Helke (1975 manuscript published in 1979). What I call zero items Helke calls virtual nouns. My symbols C and M are directly taken from his proposals for English; so is my "zero-N-but-nonzero-DET" analysis of the pronoun *je* and its analogues. Helke's justification of 'virtual nouns' and my defence of 'zero nouns', however, are quite independent of each other. Not only are the arguments dissimilar, but to some extent,

the ideas are of different scope, too. Helke claims that all English pronouns are determiners. I will show that at least some Bangla pronouns count as nouns. I am claiming that, at least in Bangla, some pronouns are determiners, some pronouns are nouns, and the rest are indeterminate, pending further research.

Helke notes that Arab grammarians postulated "hidden nouns", the same concept as his "virtual nouns" (Helke 1979: 185). Fiengo (1974) also proposes, on rather different grounds, a phonologically empty noun with grammatical content, and explicitly distinguishes it from the notion of an empty N node.

3.1.9 What's in a word: cliticizing processes

Section 2.2.6 pointed out that *Ta* does nothing to the stem that traditional grammarians think it is suffixed to. Still, there are reasons to regard *Ta* as part of the same word as the 'stem' immediately preceding it. The strongest such reason is that *Ta* takes the allomorphic shape *To* after the numeral *du* and the shape *Te* after the numerals *tin, car, aRay*, and optionally after the Determiners *ey, oy, Sey* (augmented forms of *e, o, Se*). Since vowel-harmonic processes in Bangla do not, in other cases, stray beyond single-word domains, we need to postulate a clitic fusion rule fusing *Ta* with the word preceding it at some appropriate point (we shall call this process 'cliticization', slightly modifying ordinary usage), and this rule needs independent motivation. Here comes the independent motivation.

As noted in section 2.2.6, bona fide declensional suffixes (indicating Case and Number) wreak an appreciable amount of phonological havoc. So, a declensional suffix must be part of the same word as the stem preceding it. And this is a traditional assumption. But now notice that, at the stage of syntactic derivation where NUM′ Postposing occurs, declensional suffixes indicating Case must be separate words:

(30) i. je du To lok ke
"which two item person OBJ"
'the two people whom'
ii. je du To lok er
"which two item person GEN"
'the two people whose'

(31) i. je lok du To ke
"which person two item OBJ"
'the two people whom'
ii. je lok du To r
"which person two item GEN"
'the two people whose'

We conclude that at least up to the level of shallow structure (the output of movement rules) Case suffixes are words separate from the stems they will end up being suffixed to, and that at or after the level of shallow structure a cliticizing rule attaches every Case suffix to the item immediately preceding it. This rule, it seems reasonable to assume, makes each Case suffix the ending of a word which has a noun or numeral as its 'stem' or central element (for nouns and numerals both decline directly: *lok* 'person', *loker, dOS* 'ten', *dOSer*). Given an input string

like (31-i), where the element *Ta* (here vowel-harmonized to *To*) immediately precedes the Case suffix *ke*, the general assumption fails, since *Ta* is neither a numeral nor a noun. So, we postulate that (31-i) first changes into *je lok duTo ke* and then into *je lok duToke* – a derivation which does meet the demand that Case suffixes be attached only to declinable stems.

Similar processes, which I do not discuss here, must be invoked to make the zero noun analysis of section 3.1.8 compatible with Fiengo's (1978) Affix Principle which prohibits the attaching of any affix to a phonologically null element. Presumably, *ka-ke* "who-m?" starts out as */ke/#A/#/ke/* and becomes */ke/-A/#/ke/* by virtue of some rule which attaches the zero noun A to the element /*ke*/ "who?" Only after the Objective Case suffix /*-ke/* has been cliticized on to this composite word /*ke*/*-A*, yielding /*ke/-A/-ke*, is it legitimate to apply allomorphy rules which change */ke/* to /*ka/* in the environment of a declensional suffix, deriving at last *ka-Ø-ke*.

In this perfunctory sketch of the process I have suggested that *kake* starts out as a sequence of three syntactic elements – as a DET N CASE sequence – which certain rules then agglutinate. Joanne Grumet has suggested to me that pronouns be regarded as words which the syntax cannot break into. All the grammatical processes which I have been describing then become word-internal and must take place in the morphological component(s). The elements A, C, and M become declensional suffixes. I have found no considerations which choose between Grumet's proposal and my analysis. Perhaps the availability of two distinct proposals is an artifact of the lexicalist theoretical framework in its current form and will disappear as the framework evolves.

Let us return from this digression to the discussion of (30) and (31). The word-breaks shown in *je lok duToke* are those of the standard orthography. If the derivation were being written strictly I would write */Ta/* throughout and add a fourth line to show */Ta/* becoming *To* in late morphology or early phonology.

Since analogous reasoning shows that *khana* must also encliticize to its stem (and, like *Ta*, must do it before Case suffixes attach to their stems), and since there seems to be no point base-generating some Denominators as separate words and others as part of a composite word beginning with a Numeral, let us assume that *gulo* also starts out as an independent word and then, at or after shallow structure, fuses with its left neighbour.

That leaves only the human noun and personal pronoun pluralizer *ra* starting its life as a declensional suffix rather than an independent word. On methodological grounds we should probably claim that *ra* too gets attached to its noun or pronoun by a cliticization rule, one which precedes Case suffix attachment and Denominator attachment. As the paradigms show, *ra* plus the Case suffixes (Objective and Genitive) yields *der*. Presumably this is an early morphological process.

I should stress, though, that the cliticizing processes posited are empirically motivated only for the Denominators other than *gulo* and for the Case suffixes – not for *gulo* and *ra*. It is possible that empirical reasons will be found for generating *gulo* and *ra* as suffixes. As things now stand, *gulo* is considered a Denominator with highly restrictive selectional features (which permit it only after *kOtok* 'a few', *Onek* 'many' and *SOb* 'all') and strict subcategorization features (hence its non-occurrence in the intra-NP environment –[$_{NP}$ ___NUML/Ek/N] (see (5) and (6) in chapter 2).

3.1.10 Why some pronouns are nouns

Given the above analysis of Case and Number markers, we need extra phrase structure to accommodate what are now independent words at deep structure. So, we postulate rules (32) and (33). The only Plural element will be *ra*, since *gulo* counts as a Denominator. The N slot may be filled by some overt word or by A, C, or M. These rules predict that there will be at most one *ra* (or *der*, which is *ra* + OBJ or GEN) per NP. Hence the well-formedness of (34) and the ill-formedness of (35).

(32) NP → (DET) (NUM′) N (PLURAL) (CASE)

(33) NUM′ → NUML DENOM

(34) i. tara 'they' $[_{NP}\,[_{DET}$ ta$][_{N}\ \ ][_{PL}$ ra]
ii. Sey amlara 'those bureaucrats'
iii. je amlara 'which bureaucrats'
iv. ey amlara 'these bureaucrats'
v. oy amlara 'those bureaucrats'
vi. kon amlara 'which bureaucrats?'

(35) i.*tara amlara 'they bureaucrats'
ii.*jara amlara 'who bureaucrats'
iii.*era amlara 'these people bureaucrats'
iv.*ora amlara 'those people bureaucrats'
v.*kara amlara 'who bureaucrats?'

Now consider (36). The forms in (36), though not the height of good style, are far better than those in (35).

(36) i. amra amlara 'we bureaucrats'
ii. tora amlara 'you bureaucrats'
iii. tomra amlara 'you bureaucrats'
iv. ?apnara amlara 'you bureaucrats'

To rule out (35) and allow (36), one must assign different structures to them. An obvious suggestion is that forms in (35) have the structure DET PL N PL, with an illegitimate DET PL sequence, while forms in (36) have instead the structure N PL N PL which a phrase structure rule like (37) generates:

(37) NP′ → ....N (PL) (NP)

Rule (37) predicts that *ami amla* 'I bureaucrat' and other singular forms should be out, as indeed they are. Rule (37) also predicts that, since NP′ is not NP, there should be no recursion – that *tomra bondhura amlara* 'you-PL friends bureaucrats' and the like should be ill-formed, as indeed they are. Unfortunately (37) also predicts, incorrectly, that non-pronominal nouns should be able to occur in the N slot – that, for example, *bondhura amlara* 'friends bureaucrats' should be a well-formed string. But this string is ill-formed. Presumably strict sub-categorization features on nouns and pronouns, introduced by a general redundancy rule, can

handle this fact, although a deeper explanation may be found.

The basic schema for rule (37) is due to Delorme and Dougherty (1972). These authors argue that in English personal pronouns are nouns, and that strings like *we men* have either an appositive or a complementational structure. Halitsky (1974) has argued that their complementational proposal is superior to their appositive proposal. Without claiming that the specific arguments advanced for English carry over, I am using these ideas as hypothesis schemata. Any hypothesis about Bangla will need to be justified purely on the basis of Bangla data. Rule (37) seems to be better than an analysis of (36) in terms of words like *ami* 'I' or *tumi* 'you' being determiners. So, pending further analysis (not undertaken here), we let (37) stand.

3.1.11 Elementary NP structure: an overview

Here is a summary of the decisions made so far about NP structure. The following rules have been postulated in the syntactic component. Recall that the output of (3) – in cases where the operation applies – is construed by the semantic component as definite.

(37) NP′ → N (PL) (NP)

(32) NP → (DET) (NUM′) N (PL) (CASE)

(33) NUM′ → NUML DENOM

(3) NUM′ Postposing, optional
S.D.: NUM′ – N, where NUM′ contains *Ta* or *khana* or *gulo*
S.C.: 1 2 → 0, 2 + 1

(4) NUML Deletion, obligatory
S.D.: N – NUML – DENOM, where if 3 = *gulo* then 2 = *kOtok* and otherwise 2 = *Ek*
S.C.: 1 2 3 → 1, 0, 3

The realization of PL is *ra.* It is inserted only if the N to its left is animate, preferably human.

At least after (3) and probably after (4), the following cliticization rules apply in the order given.

(38) Attach *ra* to the word preceding it

(39) Attach *Ta* or *khana* or *gulo* to the word preceding it

(40) Attach Case to the word preceding it

Allomorph choice and morphophonological processes follow cliticization.

On the other hand, semantic interpretation, which has access to the output of (3) if not to that of (4) as well, assigns a definite reading to phrases where (3) has applied. Perhaps the relevant semantic rule applies only if NUM′ contains *Ta* or *khana* or *gulo,* in which case rule (3) need not be qualified.

Among pronouns, some have been assigned to the category N (*ami tumi tuy apni*), some to DET (*je ja Se ta o e ke kon ki kono kew kichu*, and by implication *jini tini uni ini*), some have been left indeterminate. Three zero nouns have been postulated: A, C, M. It may be useful to posit a noun /*ni/* of highly restricted distribution for the forms *ini uni tini jini*, at the phonological level /eni oni teni jeni/. This noun, an honorific counterpart to A, would have to be cliticized to its determiner early enough to effect regressive vowel harmony. The assumption of a /ni/ Noun leads to an interesting suggestion about the semi-grammaticality of (36iv), ?*apnara amlara.* Our model predicts that that this form should be grammatical and mean 'you bureaucrats', if *apni* is a Noun. But if it is instead a sequence of determiner /ap/ and noun /ni/ then *apnara amlara* should be ill-formed. Perhaps the language acquisition system of Bangla speakers never makes up its mind between these alternatives – hence the ambivalence about (36-iv).

3.2 Verb phrases

3.2.1 Layers of structure in the VP

To approach the problem, consider sentence (41).

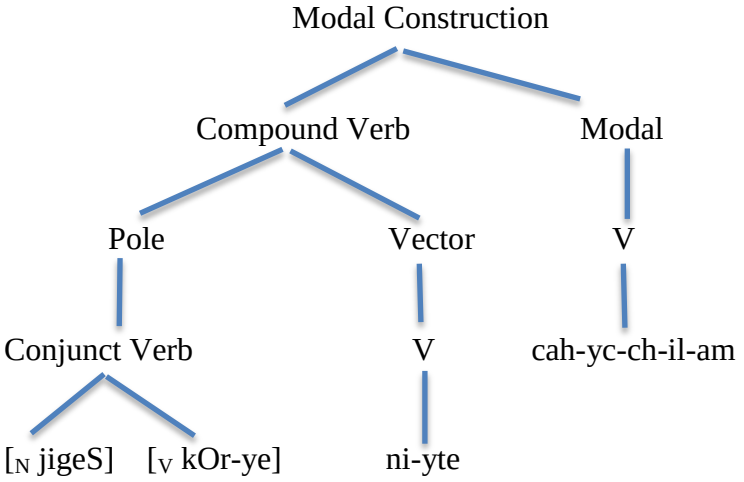
(41) i. ami jigeS kore nite cacchilam "I ask do-and to-take was-wanting"

ii. $[_S\ [_{NP'}\ [_{NP}\ [_N$ ami$]]][_N$ jigeS$][_V$ kor$][_{ENDG}$ e$][_V$ ni$][_{ENDG}$ te$][_V$ ca$][_{ENDG}$ c$][_{AUX}$ ch$][_T$ il$][_{ENDG}$ am]

We must seriously consider the possibility that (ii) might be the appropriate surface structure for (i), with cliticization absorbing *e*, *te*, and *c-ch-il-am* into V stems to form the words shown in (i). And yet there are many traditionally established facts to be accounted for which surely call for a richer syntax than (ii). The modal verb *cacchilam* governs the infinitive complement *jigeS kore nite*. In *cacchilam,* the *cac* is itself an infinitive form and *chilam* an aspect-marking auxiliary couple. In *jigeS kore nite*, *nite* is a vector verb indicating the manner of the pole or main verb *jigeS kore*. In *jigeS kore, jigeS* is a noun and the stem *jigeS kor* is a conjunct verb stem. The sequence of pole *jigeS kore* and vector *nite* is called a compound verb form. Holding the string constant, it is possible to alter the pole-vector relation between *jigeS kore* and *nite*. If instead of playing vector to the pole *jigeS kore*, the verb *nite* acts as a pole itself, then *jigeS kore nite* gets interpreted as 'to take, having asked', and the entire sentence now means 'I wanted to take (it) after asking'. If *nite* plays vector to the pole *jigeS kor, jigeS kore nite* means something like 'to ask', and the sentence means 'I wanted to ask' or 'I wanted to first ask'. A structure as slight as (41-ii) surely cannot bear the burden of so much intricate relational information.

As a first approximation to something better, let us draw a sloppy notional diagram.

(42)



The claim is not that there are syntactic categories called 'modal construction', 'compound verb', 'pole' etc. These are conventional labels for the constructions we will now study.

3.2.2 Non-verbs in verb stems: Conjunct verbs

Although it may come as a surprise to those who equate the notions 'a class of function words' and 'a closed word class', Bangla one-word verb stems are a closed class. No new members are being added to it. No mechanisms exist whereby one-word verb stems can be derived from non-verb words. (And yet it would be perverse to say that the class of one-word verb stems was a 'class of function words' in Bangla.) Verb stems consisting of two words – the first a noun or an adjective and the second a verb stem – are a large and open class. The verb stem /jigeS kOr/ 'ask' in (42) is such a two-word verb stem. All new verb stems in Bangla are two-word 'Conjunct' stems (readers unfamiliar with terminological practices in South Asia may need to be told that the writing systems of the subcontinent have what are called 'conjunct letters', characters that fuse two consonant symbols into one, and that the association is with this use of the word 'conjunct', not to the domain of coordinate conjunction): /layn kOr/ [*a piece of 1970s undergraduate slang – I observe in 2020*] 'go out with, date', /gEMn de/ 'pontificate', /awTlayn hO/ [*equally confined to the 1970s – another editorial comment added in 2020*] 'be derailed', /tEg kOr/ 'abandon', /SaMtar kaT/ 'swim', /Seddho hO/ 'boil' (in the sense 'be boiled in a liquid medium'). In some cases the first word is never found except as part of the conjunct verb: /jiges kOr/ 'ask' (no independent word *jigeS*), /Ter pa/ 'feel' (no independent word *Ter*), /ghapTi mar/ 'hide, lie low' (no independent word *ghapTi*), and to some extent /gEMn de/ 'pontificate' – the independent word *gEMn* tends to mean 'knowledge' in a matter-of-fact way, without the colourful or ironic connotation of /gEMn di/. Words like *jigeS, Ter, ghapTi* cannot be reliably classified as N or A (I call them nouns on the assumption that there is a nominal default when no other category can be established); there is never any distributional evidence for either classification. I have no principled solution to this problem. In at least one case the second word is never found except as part of a conjunct verb: /bhalo baS/ 'love'; the

only other use of *baS* as a verb stem is in the quasi-word /kharap baS/ 'unlove' (*bhalo* 'good,' *kharap* 'bad'), which people sometimes use as a foil to /bhalo baS/. Quite frequently, even if both constituents of a conjunct verb have independent use and therefore ascertainable independent meaning, the meaning of the whole has no direct relation to the sum of the meanings of the parts. Thus /SaMtar kaT/ 'swim' has nothing to do with /kaT/ 'cut'; nor does its alternative /SaMTar di/ 'swim' have anything to do with /di/ 'give'. You cannot say /SaMTar kOr/, literally "do a swim".)

I will pause in this browsing session (reminiscent of some of the less edifying enumerations of oftens and occasionallies by traditional Sanskrit grammarians) by looking at the conjunct verb stem /aRi pat/ 'eavesdrop'. The noun *aRi* is a performative word used in the phrase *tomar SOngge aRi* "you-Gen with ARI" 'ARI with you' (or *apnar SOngge aRi,* or any of the four other ways of saying 'with you', varying plurality and honorificity values); by uttering this phrase to Y, X declares silence-war – declares X's intent not to speak to Y (whether Y speaks to X or not) until X's wrath has been assuaged; ARI is a common children's game. The verb stem which means 'declare silence war' is /aRi kOr/, literally 'do ARI'. The second constituent of /aRi pat/ 'eavesdrop' – the verb stem /pat/ – can be glossed as 'spread' when used of blankets or sheets or newspapers, or as 'make' when used of yogurt. None of these meanings of /pat/, added to the meaning of /aRi/, seems to yield 'eavesdrop', and yet /aRi pat/ does mean 'eavesdrop'.

An empirical question: is each conjunct verb a single lexical entry, or a result of mutually blind or semi-blind lexical insertion into contiguous slots of an X V structure where X is N or A? In view of some of the idiosyncrasies and colligations cited above, I would suggest the first option. Every conjunct verb is a single lexical entry. But is this entry a single word, to be inserted under a simplex V node in the phrase-marker? I guess not, since the two constituents of a composite can be separated by movement rules:

(43) lolit SaSSotike bhalo na baSle
"Lolit Saswati-OBJ good not V-Protatic"
'if Lolit does not/did not love Saswati'

(44) bhalo kEno baSo?
"good why V-2p."
'Why do you love?'

I shall assume, then, that the entries for conjunct verb stems in the lexicon are of the form N V and A V (N for some conjuncts, A for others) and that they get inserted into a configuration generated by the rule

(45) U → ({N, A}) V

where the nature of U (for Unknown) is to be determined. Bar notation suggests V′. But it seems cumbersome to consider every single verb stem a 'conjunct verb without a first constituent' – to keep repeating $[_{V'}[_{V}$ in syntactic structures. So, it seems necessary to label one-word verb stems as simply V. But one-word verb stems and two-word verb stems share enough lexical and syntactic properties to be considered members of the same category. So, U should be V itself:

(46) V → ({N, A}) V

This poses two apparent problems, which disappear on closer inspection.

Problem one: recursion: [$_V$ N[$_V$ A[$_V$ N[$_V$ N V]]]] and the like. These category configurations, to which no real strings correspond, will indeed be generated by the syntax. But, since the highest node of such a configuration will always be V, the processes of semantic interpretation will refer the whole thing to the lexicon to find what it means and will find that no V stem containing three or more words exists. Thus, the misgenerated structures will be rejected as uninterpretable. Instead of treating the symptom in this fashion we might prefer to eliminate its source – wrong lexical insertion. That brings us to –

Problem two: non-conjunct insertion: lexical insertion, not 'knowing' that only ready-made N V and A V lexical entries are to be inserted, may mistakenly insert any old N or A into the first slot and any old V into the second slot, often yielding non-existing conjuncts like **SaMtar kOr* 'do a swim'. Again, however, it seems simpler to let the syntax and the lexicon jointly misgenerate and let the process of semantic interpretation seek, and fail to find, a lexical entry [$_V$ [$_N$ SaMtar][$_V$ kOr]], rather than to prohibit free insertion of V, N, A lexical items into V, N, A slots. But notice that problem one is a special case of problem two, in that, were it not for non-conjunct insertion into conjunct-internal slots, you would never get 'oversized' conjunct forms misgenerated, since there are no lexical entries of such sizes.

When the first member of a conjunct verb is a noun, that noun practically never bears a Case ending. Phrase structure rule (46) and our earlier decision to exclude the Case ending from the N constituent jointly imply that the first member of a conjunct should always be caseless. The few exceptions to this generalization which I have found are in the locative: *cokh-e pOR* "eye-Loc fall" 'come to the attention (of)', *ga-e de* "body-Loc give" 'put on', *mon-e kOr* "mind-Loc do" 'suppose, mind, remember', *mon-e pOR* "mind-Loc fall" 'come to mind (in the sense of surfacing in one's memory)', *mon-e aS* "mind-Loc come" 'surface in one's memory, occur to one', *mon-e rakh* "mind-Loc keep" 'remember', *mon-e thak* "mind-Loc stay" 'remain in one's memory', *mon-e hO* "mind-Loc happen" 'seem, occur (to)', *hat-e rakh* "hand-Loc keep" 'remain in the good books (of)' (the Bengali image is that YOU are keeping THEM under your secret control, not that they are keeping YOU in THEIR good books). One obvious way to handle these exceptions is to open the flood-gates and let them all in – to allow Noun + Case as a first constituent in conjunct verbs. Such a move would lead us to expect that there will be cases where the Objective Case and the Genitive Case also make an appearance. The facts do not fulfill this expectation. Thus, such a move would be overpredictive. Other options are to regard these locative forms as Particles (see chapter 8) or as Adjectives. For various reasons, I do not like these options in general (although surely *pichone lag* 'tease' is of the form P V). I will not discuss the reasons here; the discussion would constitute too long a digression. The upshot is that I must, in the absence of any real solution, note that there are unexplained exceptions to the regularity that the N in an N V conjunct V carries no overt case marker. The class of exceptions is small.

Since conjunct verbs are a productive class, the Bangla lexicon cannot list them, but must have conjunct formation rules deriving all but the most idiosyncratic conjunct stems from the individual N/A and V entries of their constituents. Presumably these rules are word formation rules in Aronoff's (1976) sense. The lexicon may count each conjunct stem as one word with an internal # (word boundary), leaving it to the syntactic component to treat

conjuncts as two-word structures for some purposes and as single words for other purposes.

3.2.3 Passives and their ilk: certain conjunct verbs

Does Modern Bangla have a passive? Yes and no. All grammar-conscious speakers of Bangla are used to the notion, thanks to the structure of three nationally important languages: Sanskrit, English, Hindi. The Sanskrit suffixal passive is hard for Bangla to mimic systematically, but forms like *chobiTa darun dEkhacche* "picture-item splendid looks" 'The picture looks splendid' or *ganTa Opurbo Sonacche* "song-item marvellous sounds" 'The song sounds marvellous' come close to being quasi-passive forms. (I would agree to call them quasi-passive if the phenomenon was systematic). The English passive with *be* or *get* can be mimicked by using conjunct verbs whose V constituent is *hO* 'be, happen, become' and whose first constituent is the participle-gerund form of the verb one seeks to passivize. So, one can translate 'Bread was being eaten' as *ruTi khaWa hocchilo* "bread eaten was-being", and 'The picture was being drawn' as *chobiTa aMka hocchilo* "picture-item drawn was-being". But it is misleading to consider such Bangla structures as any more passive than *chobiTa darun dEkhacche*, for two crucial features of cross-linguistically canonical passives are missing. First, the agent usually cannot be mentioned. There is no way, for instance, to expand the string *chobiTa aMka hocchilo* into an agent-specified sentence that would correspond to 'The picture was being drawn by the unregenerate abstractionist' (to express that meaning in Bangla, one would have to use the active voice). Second, the patient may be omitted: *khaWa hocchilo* and *aMka hocchilo* are complete sentences which mean, respectively, 'Eating was taking place' and 'Drawing was taking place'. Finally, the Hindi passive with the auxiliary *ja* 'go' can be mimicked by using conjunct verbs whose V is *ja* 'go' and whose first constituent, again, is the participle-gerund form of the verb to be passivized. But there is a systematic meaning difference. The Hindi sentence *tasviir khiiMcii gayii* "picture-item drawn went" 'The picture was drawn' morphologically corresponds to but does not paraphrase the Bangla sentence *chobiTa aMka gElo* "picture-item drawn went" 'It was possible to /They were able to draw the picture'. Another example: *mobil OYler bOdannotaY masTarpis thiyeTar dEkhano jacche* "Mobil Oil's generosity-LOC Masterpiece Theatre shown is-going" 'Masterpiece Theatre is being able to be shown through the generosity of Mobil Oil' (poor diction in English, but this translation needs to be as accurate as possible). Again, even if one disregards the difference between 'to be' and 'to be able to be', the agent is not (is *never*, in this case) indicated and the patient may be omitted, so the construction cannot properly be called a passive construction.

I submit that no special grammatical machinery is needed to deal with the ability of certain Bangla structures to mimic the passive constructions of foreign languages. I will discuss participle-gerunds and *dEkha* 'look' and *Sona* 'sound' in due course, in relation to the full range of functioning of these forms. Bangla has no 'passive transformation'. Nor does the structure of Bangla call for a Bresnan-type system (Bresnan 1978) of matching the subject of certain clause types systematically with the object of other clause types, since, to use the transformational metaphor, the 'object of the active sentence' does not 'become' the 'subject of the corresponding passive sentence' in the Bangla imitations of the foreign passives. Case-markings remain what they 'were' – one finds *o-ke* 'hi-m /he-r' in both *amra oke paThabo* "we him/her will-send" 'We will send him/her' and *oke paThano hObe* "him/her sent will-be" 'S/he will be sent'. Only some relational grammarians who want to find passivization everywhere

will be willing to make the definition of 'passive' so broad as to endow even Bangla with passive forms. Such moves, which relational grammarians have in fact made for other languages, strike me as somewhat similar to setting up a category of 'imperative' which covers *you*-Deletion and *please* and *why don't you* and *could you* – possibly useful in an attempt to understand how certain linguistic structures lend themselves to communicative use, but for the time being formulated in a theoretical limbo without any explicit relation to elementary points of constituent structure (e.g. the specific properties of Bangla conjunct verbs and stems like *dEkha, Sona*). However, this is a response to the present situation of relational grammar, which may change. (*Editorial note dated 2020: the situation did change; while writing this thesis I had no access to a pertinent relational grammar thesis that was being written at the same time, whose author later switched over to lexical-functional grammar. It would be inappropriate for me to emend the wording of this passage today.*)

3.2.4 Compound verb equals pole plus vector

The heading of this section lays out the functional composition of a 'Mannered' (Manner-marked) or 'Compound' verb stem. For a detailed study of compound verbs on which the following exposition draws heavily, see Dasgupta 1977.

Of the compound verb's two constituents, the vector (chosen from a special set of verbs – all of which also serve as poles) indicates the orientation or manner of the process, and the pole (which may be any V, and is in this sense 'freely' chosen) indicates the process itself.[6] A conjunctive /ye/ ending marks the pole. I will gloss V-*ye* as "V-and" in the morph by morph literal glosses. English verb-particle sequences, which themselves have been called compound verbs, often translate Bangla compound verbs rather well: *boSe pOR* "sit-and fall" 'sit down', *nece bERa* "dance-and roam" 'dance around'. Like verb-particle sequences in English, Bangla compound verbs tend to come from a restricted but reasonably numerous list, and yet seem, in the hands of talented authors and speakers, to manifest living principles of combining particular poles with particular vectors to some effect. The lexicon probably never quite needs to list any compound verb stem as an item, but in the following example it comes uncomfortably close to needing to do just that. The vector *paTha* (a stem which as a pole means 'send') only goes with the poles *bOl* 'say', *lekh* 'write', and *Dak* 'call'. One can describe this situation either by listing *bole paTha, likhe paTha, Deke paTha* as lexical entries or by forming them by means of a statement like 'the vector /paTha/ requires a verb-of-communication pole'. My guess is that the relevant statement cannot be worded accurately enough and in any case would cost more than three lexical entries would. But at least one can distinguish the alternatives. If *paTha* were to co-occur only with *bOl*, it is not clear that one would have been able to make any distinction at all between a 'lexical entry *bole paTha*' and a 'statement which constrains /paTha/ to take /bOl/ as its one and only pole.'

The principles which determine admissible pole-vector combinations are a fascinating

[6] Dasgupta (1977) states that the pole stem is freely selected. Michiko Kosaka has pointed out to me that that statement may also mean that the lexical insertion of the pole V is not based on information about what else is already in the tree. But I suppose that pole V insertion does follow and depend on vector V insertion. I have changed the wording to preclude the unintended reading noticed by Michiko Kosaka.

topic of inquiry. As Dasgupta (1977) argues, one must set up lexical affiliation fields (an only partly understood new type of formal object) in order to characterize the way in which, if vector verb stems a, b, and c used with 'less transitive' poles express Manners MA, MB, and MC, then their partners, vector verb stems a′, b′, and c′, used with 'more transitive' poles, express Manners MA, MB, and MC – where stems a, b, and c when used as poles (recall that a stem which can function as a vector can also function elsewhere as a pole) are intransitive and their partners a′, b′, and c′ when used as poles are very roughly their transitive counterparts, and where the 'less transitive'/'more transitive' relation is one that an intransitive pole bears to its transitive counterpart and a transitive pole bears to its causative counterpart. Since the patterns in an a-a′ pair do not have to be exact counterparts, it occasionally happens that some vector a has two partners, a′1 and a′2, expressing two different Manners MA1 and MA2. In these cases the vector a will be ambiguous between MA1 and MA2. Let us call such a vector bigamous and let us call a vector with only one partner monogamous. Using this imagery one must say that many vectors are spinsters. Spinsters occur freely with intransitive, transitive, and causative poles. At least one vector, *aS* (which, when used as a pole, means 'come'), is a spinster when it is associated with the meaning 'the action is gradually flowing from the past to the present', but is monogamously paired to a partner when it bears the reading 'the action is nearing completion', a reading it shares with its partner; that partner, *an*, has the reading 'bring' when it is used as a pole. Since /aS/ is perhaps the only vector that has both statuses, perhaps we need not waste our terminological ingenuity deciding to call it a 'hemigamous' vector. But there may turn out be several cases outside Bangla that will make such a term useful.

Another principle (not discussed in Dasgupta (1977)) which governs pole-vector juxtapositions is that vectors by and large require a non-identical pole. Thus, as far as other grammatical and lexical principles are concerned, *poRe pOR* "fall-and fall" 'fall down' and *beRiye bERa* "roam-and roam" 'roam around' could just as well exist. But these stems are in fact ill-formed. The only exceptions to this taboo are *diye de* "give-and give" and *niye ne* "take-and take", which occur and which have meanings best described as intensive versions of 'give' and 'take'. Ranjit Chatterji pointed out the taboo to me; I noticed the violations. Presumably an adequate response involves postulating lexical entries for *diye de, niye ne* and then setting up a principle that forbids the formation of compound verb stems with pole and vector lexically identical. That principle, governing as it does the formation of new stems rather than the right of lexically listed stems to exist, will not affect *diye de, niye ne*. This proposal takes a leaf out of Aronoff's (1976) book.

From the question 'Which poles co-occur with which vectors?' we turn now to the question 'What is the constituent structure of a compound verb?', to which also Dasgupta (1977) provides a detailed answer. The most important part of the answer is a phrase structure rule:

(47) V → (V) V

It was presumed that the pole's obligatory /ye/ suffix would be handled by auxiliary hypotheses that the paper did not have room for. Let us tentatively suppose that some lexical or transformational device (details irrelevant) introduces /ye/. This allows us to state (47) quite simply and to compare it with (46), the rule posited in section 3.2.2 for conjunct stems and repeated below (with its number unchanged):

(46) V → ({N, A}) V

A theory incorporating (46) and (47) predicts that conjunct verbs and compound verbs should behave similarly. Any dissimilarity of behaviour between conjunct verbs and compound verbs which is not explicable in non-phrase-structural terms would count against the (46)-(47) view. In order to confront this view with the data it is useful to bring in an alternative view. Let us consider, for this purpose, Ashok R. Kelkar's (personal communication, 10/10/1977) proposal that a vector verb is a Main Verb (in Chomsky's (1965) sense) and that its pole is the Head of a Verb Phrase which acts as the Complement of this Main Verb. Strict application of bar notation translates this proposal into (48). I am truncating Professor Kelkar's proposal slightly. (He proposes that the Complement to the V in (48-i) is an embedded S which gets reduced to a VP as its subject undergoes Equi-NP Deletion, i.e. deletion under identity with the subject of the matrix S. This point is independent of the (47) vs (48) issue. See chapter 5 for discussion of VP and S.)

(48) i. V′ → (VP) V – this V is the vector
ii. VP → (NP) (NP) V – this V is the pole

We will now compare (47) and (48) with respect to empirical consequences, holding (46) constant so that we are really comparing (46) + (47) with (46) + (48). The comparison will show that, where the predictions of the theories differ, (47) is correct and (48) incorrect.

In the light of the universally accepted principle that if a substring moves as a unit then it must be a single constituent, consider what predictions (47) makes that (48) contradicts and vice versa. If there is a rule moving V – and there is – then (47) predicts that a compound verb will move as a unit exactly as a conjunct verb does; or, more accurately, (47) predicts that a compound verb, being a single constituent, is entitled to such movement, not that such movement must occur. (48) predicts that a compound verb, not being a constituent, will never move as a unit. In the opposite direction, (48) permits a movement which (47) forbids. If there is a rule moving VP (and though this fact is exemplified only in later chapters, such a rule does exist), (48) predicts that a sequence consisting of the pole of a compound verb and the complements preceding it will be able to move as a unit. (47), which analyzes such a sequence as a non-constituent, predicts that such movement will be impossible.

So, we have positive and negative predictions from both theories. Let us test them.

(49), which has a conjunct verb *biSSaS koriS* that both theories regard as a constituent, exemplifies the notion 'move as a unit'. The movement rule relating (49-a) to (49-b) must treat *biSSaS koriS* as a single constituent. In (50), the sort of movement which (47) allows and (48) prohibits takes place. In (51), the sort of movement which (48) permits and (47) forbids fails to take place. Thus (50)-(51) bear out (47) and refute (48). It is irrelevant to these points whether the movement processes involved are to be considered rightward or leftward, syntactic or stylistic. In fact, this argument probably remains valid even if one posits an interpretive account instead of one in terms of movement, although I am less sure of this.

(49) a. tora or kOtha biSSaS koriS?
"you his/her word belief do"
'Do you believe what s/he says?'
b. biSSas koriS tora or kOtha?

"belief do you his/her word"
'Do you *believe* what s/he says?'

(50) a. tora or khObor jene gechiS?
"you his/her news know-and gone-have"
'Have you discovered the news about him/her?'

b. jene gechiS tora or khObor?
"know-and gone-have you his/her news"
'Have you *discovered* the news about him/her?'

(51) a. Sumit gaRiTa kine nito kintu iSita gaRiTa kine pheleche
"Shumit car-item buy-and would-take but Ishita car-item buy-and has-dropped"
'Shumit would have bought the car but Ishita has preempted him'

b. *gaRiTa kine Sumit nito kintu iSita pheleche
"car-item buy-and Shumit would-take but Ishita has-dropped"
'Shumit would have, but Ishita rashly has, bought the car'

In view of the fact that it is unclear whether the above argument is valid if one assumes interpretation instead of movement, I shall compare (47) and (48) with respect to their ability to handle certain idiom facts. In many languages, idiomatic expressions include words which do not occur independently: the French noun *fur* in *au fur et à mesure,* the English noun *midst* in *in the midst of,* etc. Consider three Bangla verb stems of this sort, /kaMc, ranga, heda/, which occur only in the compound verbs *keMce ja* 'go wrong' (said of plans), *rangiye de* 'colour, redden, brighten (tr)', *rangiye tol* 'gradually brighten (tr)', and *hediye ja* 'become bored and disgusted'. Now, a proponent of (47), which assumes that a compound verb is a V, can accommodate these facts so easily (by setting up lexical entries for *keMce ja, rangiye de, rangiye tol, hediye ja*) that (47) can almost be said to predict that this class of idioms should exist. In contrast, (48) makes it necessary to assume that there are lexical items /kaMc, ranga, heda/ and that they have selectional features to ensure correct lexical insertion. These features must be very specific: for *heda*, for example, the feature must say 'occurs as head V of a VP complement to the vector /ja/'. It is like having in French a lexical entry for the noun *fur* which says it must occur between *au* and *et à mesure*! Perhaps such assumptions are not intolerable, but it seems better to maximize the number of idioms that have unitary lexical entries, and therefore to select (47) and not (48) as an analysis of the compound verb.

A real-life proponent of (48) may, in contrast to the straw man I have just been controverting, propose that there are indeed lexical entries for *keMce ja* etc., but that they do not have the form V V (such structures, of course, do not exist if you assume (48)). Instead, they are Conjunct Verb stems with a first member of indeterminate N or A category, similar to forms like *Ter pa* "get *Ter*" 'feel' which I have discussed in section 3.2.2.

Such a view may appear unassailable, since words *keMce, rangiye,* and *hediye* do not occur outside these two-word verb stems and therefore, you might think, cannot be shown to be verbal conjunctive forms.

In response, I would point to the verb stems *kaMciye de* 'cause to go wrong' and *renge oTh* 'gradually brighten (intr.)', which bear to *keMce ja* and *rangiye tol* exact paradigmatic relations of a sort described in Dasgupta (1977). These relations become mysterious if one

assumes that *keMce, kaMciye, renge, rangiye* are not verbs: only verbs take the causative suffix. As for *hediye,* it has the wrong phonotactic shape for an A or an N. It could perhaps be a Particle; I do not feel I need to argue that it is a Verb, though I don't see how one would show that it wasn't a verb. So, the attempt to save (48) collapses. There is a stalemate over *hediye* and clear superiority for (47) in the cases of *keMce* and *rangiye.*

The above arguments selecting (47) over (48) harboured an ambiguity which served to make the formulation simpler without vitiating the arguments and which can be removed now. (Unfortunately, the following discussion is not definitive for the purpose of this volume. See chapter 11 for the final decision on this and other matters.) In what I have called V V the first "V" in my notation is ambiguous between the pole stem and the pole-/ye/ sequence; the second "V" is ambiguous between the vector stem and the vector-(auxiliary)-(tense)-ending sequence. Using *V* for stems and *V′* for verb words, let us introduce phrase structure rules like

(52) V′ → V (AUX) (TENSE) ENDING

(53) V′ → (V) V′

(54) 'sat down': $[_{V'}[_{V}$ /bOS-ye] $[_{V'}[_{V}$ [pOR]$[_{T}$ yl]$[_{E}$ am/]]] "sit-and fell"

Rules (52)-(53), generating as they do structures like (54), still hand-wave about the conjunctive /ye/ suffix, but do correctly separate the vector word and the pole word as sister constituents; each of them can prepose on its own and must stand as a distinct constituent. This constituent structure, however, renders the notion of 'a compound verb stem *boSe pOR*' problematic for the syntactic component and (on the usual assumption that a lexical entry should be a phrase-structural constituent) for the lexical component of Bangla grammar as well. One way out of the problem is to abandon the usual assumption just mentioned and to allow some lexical entries – a not fully regular minority of them – to consist of parts which are constituents themselves but do not add up to a single constituent. This option is proposed in an analysis by Emonds (1973) of English V-PP collocations where the lexical entry, which is inserted discontinuously, is not a constituent (thus, there is a lexical entry *take – to task* and a deep structure *John PAST take the students to task*, on this analysis). Sound methodology seems to me to require that this option be reserved for emergencies. Let us explore possible analyses where a single node does dominate each compound verb lexical entry.

If a single node is to dominate the lexical entry for (54), it must be a V with one or more bars – V′, say. For simplicity's sake, if some verb entries are labelled V′, then all verb entries should be so labelled. The discussion of (45) above makes this point at greater length. It is appropriate, then, to remove rules (52)-(53) from the phrase structure subcomponent of the syntax and to assign these rules instead to a new 'phrase formation subcomponent of the lexical component'. Under this proposal V′ is a lexical node which the set of syntactic phrase structure rules, in Bangla at least, rewrites no further. Substructures like (54) get generated in the lexical component, by phrase formation rules, and enter a V′ slot in an output of the (syntactic) phrase structure rules. Since rules of movement and deletion do not 'know' that some of the tree structure arises by action of syntactic rules and some of it by lexical insertion of prefabricated substructures, movement rules move and deletion rules delete particular constituents of Bangla structures regardless of whether those constituents arose from phrase structure rules of the syntax or from phrase formation rules of the lexicon.

What relation, within the lexical component, should phrase formation rules like (52) and (53) have to the substructures like (54) which they are supposed to generate? It is necessary to choose between options like (55), (56), and (57):

(55) **Step one**: Apply (53) and then (52) to V′ which is regarded as one of the lexicon's many initial symbols. Output:
[$_{V'}$[$_{V}$][$_{V'}$[$_{V}$][$_{T}$][$_{E}$]]]
Step two: Insert the rudimentary lexical entry $_{V}$/bOS $_{V}$ pOR/ into the output of step one. Output:
[$_{V'}$[$_{V}$ /bOS][$_{V'}$[$_{V}$ pOR/][$_{T}$][$_{E}$]]]
One might regard this sort of output as 'lexemic representation' in the sense that this representation shows only lexemes. The notion of Lexeme (for example, the Lexeme /dekh/ is the abstract overall characteristic of the concrete paradigm of verb words *dekhi dekhiS dEkhe* etc. in which the Lexeme /dekh/ appears) plays a central role in the Word and Paradigm approach to morphology.
Step three: Map the output of step two into a 'full' representation by introducing Tense and Ending morphemes from the small stock available, observing the idiosyncratic needs and prohibitions of particular V stems. Also, introduce the conjunctive Ending /ye/ somehow (let us not ask yet exactly how). Output:
[$_{V'}$[$_{V}$ /bOS][ye][$_{V'}$[$_{V}$ pOR][$_{T}$ yl][$_{E}$ am]]]
One might call this output 'lexetic representation' in contrast to 'lexemic'. Just as the phonological component of a grammar includes two sublevels of representation, the abstract 'phonemic' sublevel and the concrete 'phonetic' sublevel, so also one might want to distinguish two sublevels, the abstract 'lexemic' and the concrete 'lexetic', within the lexical component.

(56) **Step one**: Like step one of (55)
Step two: Introduce grammatical morphemes freely. Output:
[$_{V'}$[$_{V}$][/ye][$_{V'}$[$_{V}$][$_{T}$ yl][$_{E}$ am/]]]
Here again I hand-wave about how /ye/ comes in.
Step three: Insert the rudimentary lexical entry provided that the grammatical morphemes present in the subtree permit this:
[$_{V'}$[$_{V}$ /bOS][ye][$_{V'}$[$_{V}$ pOR][$_{T}$ yl][$_{E}$ am/]]]
This proposal too makes it possible to call the output of step three 'lexetic representation' as distinct from 'lexemic'; the difference is that 'lexemic representation' in (56) refers to the lexical entry itself and not, as in (55), to a partly filled substructure.

(57) **Step one**: Like step one of (55).
Step two: Independently, combine V, T, and E elements freely (restricted only by the $V_1 - V_2$ sequencings that lexical entries dictate) and generate lots of sequences.
Step three: Apply the output of step one as a template to the output of step two and reject all sequences that do not fit the template. For example, reject /bOS-yl-pOR-am/ on the grounds that V T V E does not parse properly. Assign constituent structure to sequences which pass.
Step four: Insert /ye/, somehow (the issue of just how is being set aside for separate study, later in the discussion).

The choice between (57) and (55)-(56) strikes me as straightforward. No empirical or theoretical considerations argue for either option. So, the choice turns on a point of method – a significant point of method, I believe. The question, as I see it, is: in a situation where a grammar generates a set of data correctly by applying *n* mechanisms in concert but misgenerates wildly if, instead, *n* – 1 mechanisms come into play, should there be reasonable limits on how wild such misgeneration is allowed to be? For example, is it methodologically okay, to echo an example of Seegmiller's (1974), to write an algorithm generating permutations of the entries in Webster's dictionary and then applying a set of templates and filters to isolate all and only the sentences of English out of this random output? To my mind, and to Seegmiller's, the answer is that it is not okay to build such grammars; that the errors which result when a small part of the grammar fails to act should be in some sense 'possible errors' – though not necessarily the sort of errors one would expect from a human learner. On these methodological grounds one should reject (57) forthwith, in the absence of any argument in its favour. That leaves us with (55) and (56). It will not take us long to select (56). Throughout Chomsky (1965), this problem type crops up repeatedly. The general question is: do lexical morphemes enter the tree first and grammatical morphemes come to flank them later, or vice versa? The general answer seems to be 'vice versa'. In the case at hand, also, one can specifically argue for this conclusion. Here is such an argument.

There is a class of verbs like /biMdh/ 'pierce', /ub/ 'evaporate' etc. which never co-occur with honorific endings. The obvious way to express this fact is to insert endings into substructures first and then to make sure that the substructures into which one inserts verbs of the /biMdh/ class contain a non-honorific ending. If one tries to do it the other way, putting the verb stem in first and then letting the appropriate endings get tagged on, then one has to find some way of classifying the verbs independently of the endings which groups /biMdh/, /ub/ etc. into a single group G so that honorific endings can be instructed not to co-occur with members of G. But the members of G seem not to share any properties except that honorific endings avoid them. So, it is better to insert the verb into a frame already endowed with the grammatical morphemes flanking the V node. Further arguments of the same form are readily available. They surely need not be marshalled one by one. Let us conclude that (56) is better than (55).

(55)-(57) all explicitly sidestepped a problem to which we now turn – that of inserting /ye/. The point of choosing (56) out of this set of options without tackling the /ye/ problem was to show that (56) is the best of those options regardless of how we deal with /ye/.

With respect to /ye/ the main options are: (i) that it is present in the rudimentary lexical entry which constitutes input to step three of (56); (ii) that a word formation rule introduces it; (iii) that the recursive phrase formation rule expanding V′ explicitly mentions it: V′ → (V /ye/) V′; (iv) that it is adjoined to V by an obligatory transformation V V′ →V+/ye/, V′; and (v) that it is freely inserted into, and semantically interpreted in, the E position in a structure $[_{V'}\ [_{V'}$ V E] V′] which reflects the correct phrase formation rule V′ → (V′) V′.

Elaine Marsh has pointed out to me that option (i) is inadmissible because to list the /ye/ would obscure its predictability. Michael Board has pointed out to me that option (ii) is inadmissible because word formation rules alter meaning appreciably and the /ye/ of compound verbs is meaningless. Independently of Board's observation, (ii) would fail for the many compound verb stems which must be listed as compound entries rather than formed by word formation rules. Option (iii) runs afoul of the fact that V+/ye/ acts as a constituent.

Options (iv) and (v) remain as the only serious contenders. Option (v) relies on a not yet formulated semantic rule which will ensure that [$_{V'}$[$_{V'}$ V E]V′] triggers access to compound verb lexical entries only when E is /ye/. Since the need for some such semantic rule may emerge in future research, this may turn out to be a point in favour of (v) rather than against it. Only detailed semantic investigation will tell. Syntactically, though, option (v) is less good than option (iv) in cases whose existence has not been acknowledged so far in this chapter – cases where the compound verb contains one pole and two vectors, e.g. *naciye niye bERaY* /naca-ye ni-ye beRa-e/ "cause-to-dance-and take-and roam-s" (approximate gloss: 'manipulates'). For such phrases options (iv) and (v) differ in that (iv) assigns structure (58) whereas (v) assigns both of the structures shown in (59) without choosing between them.

(58) [$_{V'}$[$_{V}$[$_{V}$/naca][ye]][$_{V'}$[$_{V}$ni][ye]][$_{V'}$[$_{V}$beRa][$_{T}$ Ø][$_{E}$ e/]]]]

(59) i. [$_{V'}$[$_{V'}$[$_{V}$/naca][$_{E}$ye]][$_{V'}$[$_{V'}$[$_{V}$ni][$_{E}$ye]][$_{V'}$[$_{V}$beRa][$_{T}$ Ø][$_{E}$ e/]]]]
ii. [$_{V'}$[$_{V'}$[$_{V'}$[$_{V}$/naca][$_{E}$ye]][$_{V'}$[$_{V}$ni][$_{E}$ye]]][$_{V'}$[$_{V}$beRa][$_{T}$ Ø][$_{E}$ e/]]]

Now, evidence exists that the major constituent break falls between *naciye* and *niye bERaY*. To wit, V′-Deletion deletes *niye bERaY*, as in the sentence *kukurTake haMTiye Ø nOY, naciye niye bERaY* "dog-item-OBJ cause-to-walk-and Ø not, cause-to-dance-and take-and roam-s" '(S)he doesn't take the dog for a walk, (s)he takes it for a dance', where Ø indicates the deletion site (notice that it does not matter here if the process turns out to involve interpretation and not deletion).

In view of this evidence, theory (iv), which predicts the unique and correct constituent structure (58), is superior to theory (v), which predicts the wrong structure (59-ii) as well as the wrong structure (59-i) and fails to ensure selection of the right structure. So, we conclude that the right way to introduce /ye/ in compound verbs is to use transformation (60), *ye*-Insertion:

(60) *ye*-Insertion, obligatory
S.D.: V – V′
S.C.: 1 2 ➔ 1+/ye/, 2

However, (60) poses a mechanical problem. How can we stop (60) from reapplying to its own output and generating **/boS-ye-ye pOR-yl-am/, */boS-ye-ye-ye pOR-yl-am/,* etc? The S.D. will always be satisfied, given the effect of adjunction. One can always solve this problem by reformulating (60) as a rule which assigns the feature CONJUNCTIVE to the first factor, leaving it up to a late rule to spell out the morphological content of the feature. There may be other and better solutions. In the absence of any compelling reason to pick a particular refinement, I will stick to the crude rule (60) as a provisional formulation – bearing in mind that additional machinery (or, perhaps, a slightly modified analysis) is needed to get the right results.

Notice that (60) works just as well for the relatively rare two-vector compound verbs as for ordinary, one-vector, compound verbs. There is no standard term for the more infrequent type as yet. K.D. Bose is quoted by D. Zbavitel (1970: 114) as having proposed the Bangla term *SOngjukto jowgik kriya* 'combined compound verb' for two-vector compound verbs. Neither Bose nor Zbavitel offers an analysis. Zbavitel (1970: 112) notes that the structure is rare, making it difficult to justify a specific analysis. I have nothing to add to Bose and

Zbavitel's lack of analysis as far as two-vector compound verbs are concerned.

Before leaving the subject of compound verbs I should talk about the absence of three-vector compound verbs and worse. One may seek to explain this absence simply, in terms of a limit on lexical combination processes due to constraints imposed by long-term memory, or complexly, in terms of an as yet unestablished principle to the effect that, if a recursive phrase structure process involves function words, only one recursion is permitted; I will call this principle "Ceiling".

The simple explanation is presumably easy to defend, but cannot be elaborated without getting into psycholinguistics.

The linguistic 'explanation', in terms of Ceiling, does not add anything to Ceiling. Ceiling seems to be a natural explanation for the fact that French uses the "passé surcomposé" form *J'ai eu compris* "I have had understood" for 'I had understood' in certain contexts, but never a "passé sursurcomposé" or "sursursurcomposé", i.e. **J'ai eu eu compris*, * *J'ai eu eu eu compris*, etc. However, English seems to provide counterexamples to Ceiling. One notices multiple recursion in *up from under the branch* (recursion of prepositions) and *will have been being eaten* (recursion of auxiliaries). The English evidence is inconclusive. Non-recursive analyses of the auxiliary cases exist. And *up from under the branch* may have the structure [$_{PP}$P][$_{PP}$P[$_{PP}$P NP]], which is compatible with Ceiling, instead of the structure [$_{PP}$P[$_{PP}$P[$_{PP}$P NP]]], which is not. Alternatively, one may argue that English prepositions are not function words in the relevant respect. To sum up, it is not clear whether Ceiling is or is not consistent with known data. Further research, encompassing many languages, is needed.

We turn now to modal constructions. The discussion of modals in this chapter will be less complete than that of compound verbs. Modal constructions interact with participle-gerunds, which will be discussed in chapter 4.

3.2.5 Modal constructions

In this investigation the term Modal Construction refers to a construction like *Sunte cay* /Sun-yte cah-i/ "hear-INFINITIVE want-FIRST PERSON" '(I/we) want to hear' in which the first verb, in the infinitive, serves as complement to the second, which we shall call the Modal verb with no independent justification for this descriptive label. We ask: what is the constituent structure of Modal Constructions? The considerations that follow furnish only a negative answer: Modal Constructions differ from Compound Verbs. See chapter 4 for a positive characterization.

3.2.5.1 Double negation

Given the proposals made above for the structure of compound verbs, one possibility that we must consider is that a modal construction has the same structure as a compound verb, with the infinitival functioning as a pole and the modal as a vector. If this is the case, one expects modal constructions to exhibit compound verb behaviour. Do they? The answer is no. From this point on, the discussion in this chapter will be mostly devoted to dotting the i's of this answer.

Dasgupta (1977) highlights the fact that compound verbs do not tolerate double negation (even in this language which is like English in that two negatives make a positive) –

that they tolerate at most one occurrence of the negative element *na* 'not' or *ni* 'not-PERFECT'. Thus, (61) with one negative is fine, but (62) with two negatives is ill-formed.

(61) tomar gOla Somudrer aWaj chapiye uThbe na
"your voice sea's sound drown-and will rise not"
'Your voice won't upstage the sound of the sea'

(62) *tomar gOla Somudrer aWaj na chapiye utThbe na
"your voice sea's sound not drown-and will rise not"

This 'double negation test', as Dasgupta (1977) calls it, reliably distinguishes compound verbs from superficially similar constructions of the sort discussed in Chatterji (1968) and Dasgupta (1977). Applying the double negation test to modal constructions we find that they do permit double negation and thus are not compound verbs.

(63) ami kono bondhur SOngge dEkha korte cay na
"I any friend-GEN with meeting to-do want not"
'I don't want to meet with any friend' (: I want to meet with no friend)

(64) ami kono bondhur SOngge dEkha na korte cay na
"I any friend-GEN with meeting not to-do want not"
'I don't want not to meet with any friend' (: (i) I want to meet with at least one friend,
(ii) there is no friend whom I am avoiding)

(65) o amake karur SOngge dEkha korte debe na
"s/he me anyone-GEN with meeting to-do will-give not"
'S/he won't let me meet with anyone'

(66) o amake karur SOngge dEkha na korte debe na
"s/he me anyone-GEN with meeting not to-do will-give not"
'S/he wont let me not meet with anyone' (: (i) s/he won't permit that I see no one,
(ii) s/he will force me to see everybody)

(67) ami karur SOngge kOtha bolte pelam na
"I anyone-GEN with word to-say got not"
'I didn't get to talk with anyone'

(68) ami karur SOngge kOtha na bolte pelam na
"I anyone-GEN with word not to-say got not"
'I didn't get to not talk with anyone' (: (i) I didn't get a chance to be interactionless,
(ii) I was compelled to talk with everybody)

(69) amar karur SOngge kOtha bolte holo na
"I-GEN anyone-GEN with word to-say got not"
'I didn't have to talk with anyone'

(70) amar karur SOngge kOtha na bolte holo na
"I-GEN anyone-GEN with word not to-say was not"
'I didn't have to not talk with anyone'
(: (i) it didn't become necessary to avoid all talking,
(ii) it didn't become necessary to avoid anyone in particular)

I have so far found only two modals which prohibit double negation: *lag*, *par*. They are fairly instructive exceptions.

3.2.5.2 Modals averse to double negation

3.2.5.2.1 The modal /lag/

The Modal *lag* prohibits all negation, single or double, and is peculiar in other respects as well. Unlike its Hindi-Urdu counterpart *lag* 'begin', the Bangla *lag* has a highly restricted distribution. It always takes the past tense with the suffix *yl* and means 'enter a state of —ing,' where '—' is the infinitival complement of the modal. Besides, *lag* neither takes negation itself nor permits it readily in its infinitival complement. (71) is out. (72), though bad, is slightly less intolerable than (71). Small wonder, then, that (73) should be ungrammatical.

(71) *cheleTa biSkuT khete laglo na
"boy-item biscuit to-eat LAG-PAST not"
'*The boy didn't enter a state of eating biscuits'

(72) ??cheleTa kaj na korte laglo
"boy-item work not to-do LAG-PAST"
'The boy entered a state of not working'

(73) *cheleTa biSkuT na khete laglo na
"boy-item biscuit not to-eat LAG-PAST not"
'*The boy didn't enter a state of not eating biscuits'

3.2.5.2.2 The modal /par/

This word, which means 'can or may', normally takes an infinitival complement. (One example of what it 'ab-normally' does is discussed later in this section. Another example is the sentence *ami e kaj parbo na* "I this job will-be-able not" 'I won't be able to do this job'.) It occurs negated itself, or with a negated infinitival, or neither.

(74) o mangSo cibote pare na
"s/he meat to-chew can not"
'S/he cannot chew meat'

(75) o mangSo na khete pare kintu mach dibbi khaY

"s/he meat not to-eat may but fish all-right eats"
'S/he may not eat meat, but s/he eats fish all right'

(76) o ghOnTaY kuRi mayl dowRote pare
"s/he hour-LOC twenty mile to-run can"
'S/he can run twenty miles an hour'

However, when both the modal and the infinitival are negated, the result is ill-formed if the second negation follows the modal: (77), (78). But, under certain conditions characterized in detail by Singh (1976), the negative marker *na* precedes rather than follows its verb. When the negative precedes *par*, double negation with the infinitive is fine.

(77) *o na haSte parlo na
"s/he not to-laugh could not"
'S/he couldn't not laugh'

(78) *madar Teresa artoSeba na korte paren na
"Mother Teresa distressed-service not to-do can not"
'Mother Teresa can't not serve the distressed'

(79) na haSte na parar kono karon ney
"not to-laugh not being-able-GEN any reason isn't"
'There is no reason for not being able not to laugh'

(80) jodi uni artoSeba na korte na paren
"if s/he distressed-service not to-do not can"
'if s/he can't not serve the distressed'

While the ill-formedness of (77)-(78) is just puzzling at first, it becomes downright mysterious in view of the grammaticality of (79)-(80). The key to the solution lies in what Aronoff 1976 calls 'blocking', whereby, for example, the existence in English of the words *fury, variety, grace* "blocks" the formation of the expected forms **furiosity, *variosity, *graciosity* from *furious, various, gracious* (cf. the well-formed *curiosity, speciosity*, which come through because they are not blocked by words like **cury, *speciety*). Aronoff (1976: 43) defines blocking as 'the non-occurrence of one form due to the simple existence of another', without formulating a mechanism whereby blocking can be implemented. In this study I follow him in using the notion without committing myself to any mechanism for its implementation.

(77) and (78), it turns out, are blocked by (81) and (82).

(81) o na heSe parlo na
"s/he not laugh-and could not"
'S/he couldn't help laughing'

(82) madar Teresa artoSeba na kore paren na
"Mother Teresa distressed-service not do-and can not"
'Mother Teresa can't help serving the distressed'

The following detailed study of the construction exemplified in (81) will at first look like a digression. But, when I finish, it will be clear that the details about (81) and (82) will help answer the question as to why (77)-(78) are bad while (79)-(80) are okay.

I begin by giving the construction *na V-ye par-… na* of (81) and (82) a name – the /ye-par/ Construction – and asking how it arises.

Morphologically, the /ye-par/ Construction looks like a doubly negated compound verb. But that cannot be the case, since on the one hand this construction is doubly negated (which compound verbs never are) and since on the other hand compound verbs can occur with one negation or without negation (which, as the following examples show, the /ye-par/ Construction cannot).

(83) *o heSe parlo na
"s/he laugh-and could not"

(84) *o na heSe parlo
"s/he not laugh-and could"

(85) *o heSe parlo
"s/he laugh-and could"

(86) *madar Teresa artoSeba kore paren na
"Mother Teresa distressed-service do-and can not"

(87) *madar Teresa artoSeba na kore paren
"Mother Teresa distressed-service not do-and can"

(88) *madar Teresa artoSeba kore paren
"Mother Teresa distressed-service do-and can"

What is the /ye-par/ Construction if not a compound verb? Two possibilities come to mind. One is that the construction derives by deletion of *thakte* in (89) and (90). The other is that there is a special, idiomatic lexical entry which makes the construction's base-generation possible.

(89) o na heSe thakte parlo na
"s/he not laugh-and to-stay could not"
'S/he could not keep from laughing'

(90) madar Teresa artoSeba na kore thakte paren na
"Mother Teresa distressed-service not do-and to-stay can not"
'Mother Teresa cannot keep from serving the distressed'

Despite the appeal of the first possibility – since (89) and (90) are well-formed sentences in their own right – further consideration lends credence to the second possibility. Corresponding to (81) and (82) we find (91) which (cf. (92)) exhibits the same behaviour as (81) and (82). Since (93) shows that the deletion explanation fails to account for (91), and since presumably

the same device should explain (81)-(82) and (91), it seems necessary to base-generate the /ye-par/ Construction instead of deriving it by a deletion process.

(91) eTa upen na hoYe jaY na
"this Upen not be-and goes not"
'This cannot but be Upen'

(92) a. *eTa upen na hoYe jaY
"this Upen not be-and goes"
b. *eTa upen hoYe jaY na
"this Upen be-and goes not"
c. *eTa upen hoYe jaY
"this Upen be-and goes"

(93) a. *eTa upen na hoYe thakte jaY na
"this Upen not be-and to-stay goes not"
b. *eTa upen na hoYe *V*-te jaY na,
where *V* is some verb other than *thak*
"this Upen not be-and to-*V* goes not"

Let us then decide, pending further study, to posit lexical entries of the form $[_V [_V [_{NEG}$ na$][_V$ X$]][_V$ Y$]]$, where Y is *par* or *ja* and X is one of the many verb stems that can occur in this position. These entries are to be inserted into compound verb constituent structures subject to the restrictions exemplified above – with a negative marker following the structure and, if Y is *ja*, with a Present ending for *ja*. Notice that the substructure $[_V [_{NEG}$ na$][_V$ X$]]$ as a whole is inserted at the pole slot of the compound verb subtree, thus avoiding the problem that phrase structure rules have made no place where the negative marker could be inserted in the normal fashion. This property of the relevant phrase structure rules expresses the fact that normal compound verbs are not subject to double negation.

This concludes our discussion of the /ye-par/ construction, which we had held responsible for the 'blocking' of (77)-(78). We now ask whether, as promised, this account of the /ye-par/ construction helps us to understand the contrast between the goodness of (79)-(80) and the badness of (77)-(78).

It does. (77)-(78) are blocked by (81)-(82) because (81)-(82) 'exist' (are made possible by the lexical entries described above, given the existing conventions for lexical insertion). The reason why (79) and (80) are grammatical is that they fail to be blocked by (94)-(95), since (94)-(95) do not 'exist' (are ungrammatical).

(94) *na heSe na parar kono karon ney
"not laugh-and not being-able-GEN any reason isn't"

(95) *jodi uni artoSeba na kore na paren
"if s/he distressed-service not do-and not is-able"

The reason for the ill-formendness of (94)-(95) is that the lexical insertion process which would have generated them fails to go through. As mentioned in the discussion above, for a lexical

entry of the form $[_V [_V$Neg V] V] to be inserted, there must be a negative marker following the second V. In (94)-(95), this condition is not met. There is no negative marker following the second V. Besides, it is a general constraint on lexical insertion that unless a lexical entry is marked as discontinuous it must not be inserted discontinuously; to form (94)-(95) the insertion process would have had to violate this constraint by inserting a lexical entry partly to the left and partly to the right of the NEG node which shows up as the second *na*. For these reasons, it is impossible to generate (94)-(95), and thus impossible to block (79)-(80), which are, therefore, predicted to be well-formed.

Thus, close examination of the /ye-par/ construction has dispelled the mystery about the partial intolerance which the modal *par* exhibits towards double negation. This partial intolerance is now seen to be due to interference from a highly idiosyncratic construction which obscures the normal properties of a modal construction with the modal *par* and an infinitival complement.

The normal properties of a modal construction, to recapitulate, include tolerance of double negation, a property which distinguishes a modal-infinitival sequence from a compound verb. The discussion in sections 3.2.5.1 and 3.2.5.2 has thoroughly established that modal constructions must at some stage of the derivation be assigned a constituent structure which distinguishes them from compound verbs.

In order to determine this constituent structure, we must venture further afield. We will do so in chapter 4.

This seems to be the best juncture at which to point out that the entire discussion of negation has been overspecific in that what I have said of negation is also true of 'negative polarity' in general. Thus, compound verbs are also intolerant of negation of the pole plus interrogation of the vector. Cf. **tomar gOla Somudrer awaj na chapiye uThbe ki?* "your voice sea's sound not drown-and will-rise whether", where *na* negates the pole *chapiye* and the complementizer *ki* interrogates the vector *uThbe*. As yet, no firm knowledge exists of the exact scope of 'negative polarity' (as distinct from grammatical negation proper) even in the better studied languages, and certainly not in Bangla. So, I have restricted my attention to negation per se. Future work may redress this imbalance.

Appendix

Empirical arguments supporting ZIV against ENV

As noted in section 3.1.8, we are trying to choose between the Empty Node View ENV and the Zero Item View ZIV for phonologically null nouns. In this Appendix, I offer an empirical argument or two in favour of ZIV. I begin by repeating the formulation given in section 3.1.8 for ready reference. I wrote:

"It is time to define 'empty N' now – to specify the exact content of what was represented as $[_N\ ]$ in the derivations. Two definitions come to mind. The choice between them proves to be an empirical one.

"*ENV, the Empty Node View*: Lexical insertion rules are optional. Where the insertion process exercises its option of not applying and leaves a node dominating no terminal material, that slot is conventionally filled by the so-called concatenational element often written *e* for empty. The

empty noun of (15) is such an *e* dominated by N.

"*ZIV, the Zero Item View*: One or more lexical items exist which, though phonologically blank (hence representable as Ø), do have syntactic and semantic features. The empty noun of (15) is such a zero noun."

The passage quoted mentions (15); here, for reference, is (15) again:

"(15) The U-is-DET Hypothesis: NP → (DET) (NUM′) N, where N may be 'empty' in some sense"

In order to settle the issue, we have to look at examples that invoke rules (1)-(4) given at the beginning of the chapter; the reader is invited to look them up. The specific examples that prove crucial are (A-2a-f), which are null noun counterparts to (A-1a-f) built around the overt noun /boy/ 'book' (we are affixing "A" to example numbers here to separate the Appendix examples from those in the main chapter). What needs to be addressed is the ungrammaticality of (A-2f), in contrast to the grammaticality of (A-1f). This excursus takes the discussion into a grammatical region not explored in the main text – noun phrases including a genitive specifier; I assure the reader that it is safe to go there and that no extraneous issues peculiar to that region undermine the discussion:

A-1. a. amar Ek Ta boy "my one TA book" 'one of my books'
b. amar boy Ek Ta "my book one TA" (step derived by NUM′ Postposing)
c. amar boy Ta "my book TA" 'my book' (output derived by *Ek*-Deletion)
d. amar Ek khana boy "my one KHANA book" 'one of my books'
e. amar boy Ek khana "my book one KHANA" (step derived by NUM′ Postposing)
f. amar boy khana "my book KHANA" (output derived by *Ek*-Deletion)

A-2. a. amar Ek Ta [$_N$] "my one TA *null*" 'one of mine'
b. amar [$_N$] Ek Ta "my *null* one TA" (step derived by NUM′ Postposing)
c. amar [$_N$] Ta "my *null* TA" 'mine' (output derived by *Ek*-Deletion)
d. amar Ek khana [$_N$] "my one KHANA *null*" 'one of mine'
e. amar [$_N$] Ek khana "my *null* one KHANA" (step derived by NUM′ Postposing)
f. *amar [$_N$] khana "my *null* KHANA" (output that *Ek*-Deletion should yield)

If one assumes ENV alone, without any auxiliary assumptions, it is impossible to distinguish the ill-formed (A-2f) from the well-formed (A-1f) featuring an overt noun /boy/ 'book' and the marked denominator /khana/ – as well as from (A-2c) which features the null noun and the default denominator /Ta/ and from (A-1c) featuring the overt noun /boy/ 'book' and /Ta/.

In contrast, ZIV assumptions make the task perfectly straightforward. All we have to do is ensure that, in our analysis, the features of the marked denominator /khana/ and the zero item *C* fail to match. This mismatch has to do with some feature that the marked denominator /khana/ does not share with the default denominator /Ta/; for our purposes we can use the ad hoc label [+Piece] for this feature specification. The mechanisms that help us to account for the narrow distribution of the marked /khana/ over against the wider distribution of the default

denominator /Ta/ must involve insisting that the associated N be specified [+Piece]. A zero item like *C* originates with a fixed and minimal feature matrix that obviously does not include [+Piece]; thus, the ZIV hypothesis makes perfect sense of the ill-formedness of (A-2f) in contrast to the well-formedness of (A-2c, A-1f, A-1c). In contrast, a contentless empty N node does not have features, a fact that leaves the ENV account unable to star (A-2f).

However, more needs to be said to handle examples where an overt numeral appears in the definite output of NUM′ Postposing, such as (A-3d). We must also note that (A-2d) is fine as an indefinite surface form and that here, too, the compatibility between the marked denominator and this null noun, whatever its conceptual status, must be described adequately:

A-3. a. amar du khana boy "my two KHANA book" 'two of my books'
b. amar boy du khana "my book two KHANA" 'my two books' (output derived by NUM′ Postposing)
c. amar du khana [$_N$] "my two KHANA *null*" 'one of my books'
d. amar [$_N$] du khana "my *null* two KHANA" 'the two that are mine' (output derived by NUM′ Postposing)

We may grant that /boy/ in the vicinity of /khana/ carries a [+Piece] feature. But surely whatever factors star (A-2f) should also star (A-2d) and (A-3d), should they not? Exactly what makes the null noun in these two well-formed examples compatible with the marked denominator /khana/, given the lesson we seek to draw from the ill-formedness of (A-2f)?

My answer is that the overt numeral can carry the [+Piece] feature and must carry it to meet /khana/'s selectional requirements. There is variation across numerals with respect to the ability to bear this feature value. The larger the number, the harder it is to visualize so many pieces. As a consequence, /SatSo khana mach/ 'seven hundred pieces of fish' with the marked /khana/ is far less acceptable than /SatSo Ta mach/ 'seven hundred fishes' with the default denominator /Ta/. By the time we reach two thousand, the difference is sharp: /du hajar Ta mach/ 'two thousand fishes' is fine, but ???/du hajar khana mach/ 'two thousand pieces of fish' can be parsed only under duress, by analogy to well-formed cases, and is never liable to be said in a real-life context. Exactly how to formalize this is not immediately obvious. One solution is to claim that the pragmatics directly controls the presence or absence of the [+Piece] feature value; another is to let the grammar leave this feature optional, with the pragmatics vetoing all the instances of mismatch between numeral and denominator; mixed solutions can be devised. I leave the details open.

However, the logic of the argument requires me to admit that, regardless of how the details are specified on that score, this account of the well-formedness of (A-2d, A-3d) is neutral between the competing hypotheses ZIV and ENV. The point of providing this analysis of those two examples is only to rebut an obvious objection to my proposal regarding (A-2f).

It is possible that some readers will find unconvincing any argument that places a heavy burden on the handling of phenomena that invoke peripheral mechanisms and exceptional features. I now turn, therefore, to some evidence favouring ZIV against ENV that attracts core grammatical considerations. On the way to this evidence I must first supplement the account given in the main text, by observing that rules (1)-(4) only work for definite constructions built around a count noun. For mass nouns – which also allow the configuration *N Ta* to mean 'the N', but which are not compatible with the numeral /Ek/ 'one' or any other numeral – it is necessary to derive the definite output *N Ta* from /khanik Ta N/ 'some N, a little bit of N' via

NUM′ Postposing and *khanik*-Deletion; /jOlTa/ ‘the water’ is a case in point. With this supplement in place, I now present the evidence in favour of ZIV.

The range of data involving overt nouns and the denominators /Ta/ and /gulo/ include human examples such as /cheleTa/ ‘the boy’ and /meYegulo/ ‘the girls’. However, when these denominators are used with a null noun, only those examples are grammatical which involve a null noun with a *C* (inanimate count) or *M* (inanimate mass) reading; an *A* (animate human) reading is excluded. We see this in the case of demonstratives at (A-4) and in the case of possessives at (A-5):

(A-4) a. eTa ‘this item/ this material’: e.g. ‘this pen, this milk’, not *‘this girl’
b. oTa ‘that item/ that material’: ditto
c. egulo ‘these items’: e.g. ‘these pens, these boxes’, not *‘these girls’
d. ogulo ‘those items’: ditto

(A-5) a. amarTa ‘my item/ my material’: e.g. ‘my pen, my milk’, not *‘my aunt’
b. tomarTa ‘your item/ your material’: ditto
c. amargulo ‘my items’: e.g. ‘my pens, my boxes’, not *‘my aunts’
d. tomargulo ‘your items’: ditto

This argument is stronger than the argument from (A-2f) in that there is no overt numeral twist in this case. The facts at (A-4, A-5) are directly inconsistent with ENV assumptions, and count as evidence in favour of ZIV.

If this is how the language handles *C* and *M*, what about the demonstratives and *A*? A natural question. Bangla fuses each demonstrative DET and the zero item *A* into a pronoun word. The full paradigm of pronouns of this type was presented in the ‘they’ rows of the personal pronoun chart in section 2.2.1. One is strongly tempted to claim that these forms ‘block’ the otherwise expected DET-*A* sequences. However, that claim would have looked more plausible if the same phenomenon had been observed in the case of possessives like (A-5). But it is not. There is no non-clumsy way, in Bangla, to say ‘mine’ or ‘yours’ in the sense of ‘my aunt’ or ‘your aunt’; and the plurals one would have expected (the possessive counterparts to the demonstrative human plurals /era/ ‘these people’, /ora/ ‘those people’ and so on), namely */amarra/ ‘mine (Plural)’, */tomarra/ ‘yours (Plural)’ etc., are sharply ill-formed.

There is some unclarity in a similar domain in English usage as well. While the bare determiners *this* and *that* used as complete noun phrases (presumably with a null noun) only refer to mass and inanimate count referents, *those* (as in *those who*) frequently refers to a set of persons, an anomaly for which current accounts of the English noun phrase have no explanation that I’m aware of. Possessive pronouns like *mine*, *yours* and so on seem to be neutral with regard to the denotatum of the covert noun, but as I note elsewhere in this thesis *its* (as well as the inanimate plural *theirs*) is mysteriously unable to serve as a possessive pronoun of this type in the first place, which means that *its* and inanimate *theirs* are incompatible with either animate or inanimate empty nouns. I do not doubt that these problems will prove tractable to future researchers. For our purposes, the simplest thing is to set them aside. Our demonstration is over, and satisfies our needs. *Caveat lector*.

Editorial note: *This version of the appendix, specially written for the 2020 ‘edition’ of this thesis, replaces the appendix defended in 1979 and included in the final 1980 text of my thesis.*

*The rewriting is my response to the fact that my 1980 appendix rested on a shaky empirical foundation. The key ingredient in that foundation was the statement that the relative */je khana/ (and the demonstratives */e khana/, */o khana/) were ill-formed. It turned out that these combinations were fine in standard spoken Bangla in the early twentieth century. That they have been disappearing (although their counterparts containing an overt noun continue to thrive), and that my dialect does not have them, is a matter meriting diachronic and sociolinguistic investigation. I cannot afford to base an important theoretical point on shifty sands.*

*The judgments about */je khana/, */e khana/ and */o khana/ were at least robust for my dialect. But I made a move that rendered the argumentation completely unsustainable. Namely, working in isolation from the speech community in 1979, I convinced myself that even sequences containing an overt numeral such as /je du khana/ were unacceptable to my ear. The impression that such sequences were ill-formed for me – an impression which I had taken seriously and on which I had based the argumentation in my 1980 appendix – turned out to be an artifact of working under high pressure away from other speakers. That impression disappeared the moment I returned to India, resumed my interaction with the speech community, and realized, to my chagrin, what a shaky foundation I had built that appendix on.*

Since autumn 1980, then, I have known that this indefensible appendix needs to be replaced. But I did not at that stage intend to publish the entire dissertation. I therefore took it that I could simply drop chapter 3 and related elementary material when I "wrote the book". Only in 1995 did I abandon the idea of the "relative clauses book"; I hereby thank Mark Aronoff for having offered, as series editor, to place the book in a series he was editing for SUNY Press, and for cheerfully accepting my silent decision to drop out of the project after delivering some chapter drafts he found promising. Persistent requests from users of the dissertation have persuaded me now to make the entire text of my thesis available for public inspection. Acceding to these wishes obviously means fixing the 1980 appendix to chapter 3 as well. I assure all readers that my 2020 appendix is based on empirical material and conceptual resources that were well within the range of my 1979 awareness. Had I realized in 1979 that the appendix argument was not viable, it is likely that I would have arrived at a text approximately similar to what I have now written. I hope at least to have steered clear of anachronism: I have made no use of machinery developed after 1979 or of data that was out of my reach in 1979.

Chapter 4

OVERALL BANGLA GRAMMAR: INFINITIVES AND PARTICIPLE-GERUNDS

4.0 Strategy and tactics

Having shown that the structure of modal constructions differs from that of compound verbs, we go on to ask what structure modal constructions do have. Since the complement verb in a modal construction is in the infinitive, the answer to our question must consider the role of the infinitive. In Bangla, infinitive forms bear a special paradigmatic relation to participle-gerund forms. This relation must figure in a characterization of either infinitives or participle-gerunds. Participle-gerunds 'block' (in Aronoff's (1976) sense) infinitives in subject position. We need a grammar of infinitives and participle-gerunds which expresses this 'blocking' relation.

Accordingly, this chapter first brings the concept of blocking into sharper focus by looking at a clear case of blocking, from English, as it happens. Armed with a clear concept of blocking we will look at data which give us an idea of the blocking relation between infinitives and participle-gerunds in Bangla. An analysis of participle-gerunds follows. Finally, the discussion returns to infinitives and their place in the structure of modal constructions.

The little-studied area of English structure which section 4.1 examines in connection with blocking is also one that will figure in an entirely different line of argumentation in chapter 9.

4.1 English *both* and Blocking

Aronoff (1976: 43) defines 'blocking' as 'the non-occurrence of one form due to the simple existence of another'. Some of the clearest cases of blocking known to me, not yet pointed out (I believe) in the generative literature, involve the English word *both*, which most obviously blocks *all two* (cf. (1)-(3)) and also blocks *all* in sentences like (4)-(6) with a tacit 'two' in them.

(1) All three men were armed

(2) *All two men were armed

(3) Both men were armed

(4) Sally, Hal, and Bev all like it

(5) *Sally and Bev all like it

(6) Sally and Bev both like it

The tacit 'two' need not inhere in the sentence itself, which indicates that blocking goes beyond sentence grammar.

(7) *They all like it*: well-formed only if *they* refers to more than two people

(8) *They both like it*: well-formed only if *they* refers to just two people

Still on the theme that the phenomenon of blocking goes beyond sentence grammar, notice that no grammatical mechanism yet proposed can express the fact that (9) requires the three political groups to be semantically three distinct conjuncts whereas (10) requires two rather than three distinct conjuncts.

(9) The liberals and the socialists and the communists all like the proposal

(10) The liberals and the socialists and the communists both like the proposal

Although sentence grammar cannot, perhaps, fully describe the way in which *both* blocks what it blocks, grammarians must nevertheless accept some of the blocking-related facts as crucial data for purely grammatical purposes. For example, someone trying to distinguish *every* from *each* cannot ignore the fact that *every* patterns with *all* in that they (unlike *each*) get blocked by *both*. Moreover, grammarians must try in any case to relate *both*'s blocking of *all* and *every* to *neither*'s blocking of *none*. Ellen Cohen brought cases like (15) and (16) to my attention in 1979.

(11) Each sex has its own chauvinism

(12) ?Every sex has its own chauvinism

(13) ?All sexes have their own chauvinism

(14) Both sexes have their own chauvinism

(15) Neither of the hemispheres is self-sufficient

(16) ?None of the hemispheres is self-sufficient

On the other hand, let us note that a grammarian may *not* infer 'A and B belong to the same category' on the basis of situations where A blocks B or vice versa. Thus, there are limits to how much can be learnt from blocking. To see this, consider the fact that *neither* blocks *none*, as in (15) and (16). But *neither* occurs as a discontinuous conjunction fragment in sentences like *Neither the west nor the east is self-sufficient* and as some sort of adverb in sentences like *Neither did they like to be pushed around*; *none* never occurs in either of these roles and thus must belong to a different category from *neither*. More accurately, at least some uses of *neither* must belong to a category to which no use of *none* belongs. Thus, someone who had inferred from (15)-(16) that *neither* and *none* had the same categorical feature composition would be wrong.

4.2 Blocking and Bangla infinitives

Perhaps this more than grammatical and yet grammatically relevant notion of blocking will help in the analysis of Bangla infinitives.

(17)-(18) make it seem that in subject position infinitives are blocked by participle-gerunds. But (19)-(20) make it look like as though infinitives in complement position were blocking participle-gerunds.

(17) ja-Wa SOmbhOb hO-be na
"go-ing possible be-FUT not"
'To go will not be possible'

(18) *je-te SOmbhOb hO-be na
"go-to possible be-FUT not"

(19) amra je-te ca-ybo na
"we go-to want-FUT not"
'We will not want to go'

(20) *amra ja-Wa ca-ybo na
"we go-ing want-FUT not"

As we go further afield, the data will compel us to give up the pure idea that (17)-(20) exemplify full-scale blocking; we will, specifically, need to dilute the idea for (19)-(20). But the pure idea can serve as an initial hypothesis with which to approach the study of infinitives and participle-gerunds in Bangla. Some version of this initial hypothesis often occurs to learners of Bangla in whose first language (French, English, Portuguese, Hindi) a 'dictionary entry form' for verbs appears in the *to write* position of both (17) and (19).

Note that, in Hindi, the 'dictionary entry form' in *-naa* deserves to be called the infinitive-gerund form since it combines infinitive features with gerund features much as the Bangla participle-gerund in *-Wa/no* combines participle features with gerund features. When adequate analyses of non-finite verb forms in most modern Indic become available, South Asianists (including students of non-Indic languages in the area) will need to evaluate the appropriateness of using western terms like Participle, Gerund, and Infinitive at all in describing these languages. In this connection one may recall that in Classical Sanskrit, as in Bangla, the infinitive in *-tum* never functioned as a subject, and that the several Vedic infinitives (in *-tum, -tave,* etc.), morphologically non-nominative Case forms of verbal nouns (nouns which were derived by adding a nominalizing suffix to a verb stem and which had a defective declension in Vedic – lacked the nominative and some other Cases), had a good synchronic reason for never occurring in subject position.

Returning to the main line of work in this chapter, we will now look at the participle-gerund form in Bangla, bearing in mind our initial hypothesis that this form blocks the infinitive in subject position and is blocked by the infinitive in object position. The next four sections, 4.3 to 4.6, offer a description of the participle-gerund which pays no explicit attention to the relation between this form and the infinitive. Section 4.7 returns to this relation.

4.3 The participle-gerund's three forms and two functions

The participle-gerund form in *-Wa/no* combines a past participle function with a gerund function. Let us look at some morphic details of the participle-gerund ending.

The ending has the form /Wa/ when attached to a monosyllabic verb stem and the form /no/ when attached to a polysyllabic verb stem. After stems ending in a non-glide consonant, moreover, the alternant /Wa/ reduces to surface *-a* by virtue of the Onglide Drop rule of Chapter 2.

Note that no aspect of the three-allomorph alternation described has anything to do with whether the form is functioning as a past participle or as a gerund. Udaya Narayana Singh has pointed out to me that the same holds for the *@ / y@* alternation in the form of the Hindi Past Tense or Past Participle ending, which observes roughly the same phonological conditioning and which combines a finite with a non-finite function (the glyph @ is a language-specific abbreviation for /aa ~ e ~ii/). The following examples show the independence of the formal alternation of the Bangla participle-gerund ending from its functional diversity.

(21) apnar ciThi lekh*a*
"your letter writing"
'your writing letters'

(22) OSimer gaRi cala*no*
"Asim's car driving"
'Asim's driving a car'

(23) eSar khOborTa de*Wa*
"Esha's news-item giving"
'Esha's giving the news'

(24) apnar lekh*a* ciThi
"your written letter"
'a letter written by you'

(25) OSimer cala*no* gaRi
"Asim's driven car"
'a car driven by Asim'

(26) eSar de*Wa* khOborTa
"Esha's given news-item"
'the news given by Esha'

In the above examples, each of the three allomorphs of the ending – *a, no, Wa* – appears in both functions, the gerund function in (21)-(23) and the participle function in (24)-(26).

4.4 Relating the functions: two hypotheses

Two of the simplest ways of interrelating the two functions of the participle-gerund form will be considered here, stated as follows.

(27) The Transformational Hypothesis (TH): A gerund construction as well as a participle construction is a clause whose main verb is the gerund or participle. (21)-(23) are base-generated with and keep the structure $[_S\ NP_i\ NP_j\ V]$. (24)-(26) have the surface structure $[_{NP}[_S\ NP_i\ V]\ NP_j]$ derived from $[_{NP}[_S\ NP_i\ NP_j\ V]\ NP_j]$ by an Equi-NP rule deleting the first NP_j.

(28) The Base-Generation Hypothesis (BGH): A gerund construction is an NP with the gerund as its head N. A participle construction is an AP with the participle as its head A. (21)-(23) are base-generated with and keep the structure $[_{NP}\ NP_i\ NP_j\ N]$. (24)-(26) are base-generated with and keep the structure $[_{NP}[_{AP}\ NP_i\ A]\ NP_j]$. The A head of the AP is derived by a lexical rule from the $[_N\ V\ +\ Wa/no]$ head of the NP to which it corresponds.

The Transformational Hypothesis TH uses transformation (29) to derive (31) from (30). The Base-Generation Hypothesis BGH uses lexical rule (32) to interrelate the N in (33) and the A in (34), where the base generates both (33) and (34). Rule (32) in notation devised by Wasow (1977: 355), states that members of some class of gerunds can be turned into participles by adding a null affix and that the direct object of the gerund corresponds semantically and hence selectionally to the head noun (the 'subject') of the adjective. Wasow borrows his use of I and II for Subject and Direct Object, respectively, from relational grammar. Anderson's (1977) comments on his paper include a suggestion that thematic relations (Theme, Agent, etc.) be used instead of grammatical relations like Subject and Direct Object. I disregard Anderson's suggestion here, since grammatical relations are easier to handle in a first approximation.

(29) Equi-NP Deletion, obligatory
S.D.: $[_{NP}[_S\ X - NP_j\ - V]] - NP_j$
S.C.: 1 2 3 4 ➔ 1, 0, 3, 4

(30) $[_{NP}[_S[_{NP}$ apnar$][_{NP}$ ciThi$][_V$ lekha$]][_{NP}$ ciThi$]]$
"you-GEN letter write-ENDING letter"

(31) $[_{NP}[_S[_{NP}$ apnar$][_V$ lekha$]][_{NP}$ ciThi$]]$
"you-GEN write-ENDING letter"

(32) Gerund/Participle Rule: $\emptyset_{[[_N\ V\ +\ Wa/no]]_A}$ (II) = I

(33) $[_{NP}[_{AP}[_{NP}[_{NP}$apnar$][_A[_N\ [_V$ lekh]a$][\emptyset]]][_N$ ciThi$]]$
"you-GEN write-ENDING letter"

(34) $[_{NP}[_{NP}$ apnar$]\ [_{NP}$ciThi$][_N\ [_V$lekh]a$]]$
"you-GEN letter write-ENDING"

The following section will consider relevant evidence and conclude in favour of BGH, against TH.

One should view the choice between these alternatives in a broad perspective. BGH, by using a lexical rule rather than a syntactic transformation, suggests (correctly, it turns out) that only some gerunds can also function as participles. Evidence which directly supports BGH thus also provides indirect support for a theory of grammar which formally distinguishes partial processes (registered as lexical rules) from systematic processes (expressed as non-lexical grammatical rules).

4.5 Arguments for the base generation hypothesis

4.5.1 Argument one

A -*Wa/no* word functioning as a gerund can take the endings of the Objective, Genitive, and Locative Cases or be governed by Particles like *SOtteW* 'despite'.

(35) a. apnar ciThi *lekhar* karon
"your letter writing-Gen reason"
'the reason for your writing the letter'
b. OSimer gaRi *calanoke* prithibir OSTom aScorjo bolbe ki?
"Oshim's car driving-OBJ world's eighth wonder you'll-call?
'Will you call Oshim's driving of a car the eighth wonder of the world?'
c. khOborTa *deWate*
"the news giving-LOC"
'on giving the news'
d. khOborTa deWa SOtteW
"the news giving despite"
'despite the news being given'

The behaviour of the -*Wa/no* form thus markedly differs from that of bona fide main verbs. A bona fide main verb of a clause never bears Case suffixes or is governed by a participle. Nouns do and are. So, in the above respects, the -*Wa/no* form in its gerund role is a noun. This fact militates against TH (which regards the gerund just as a verb, not at all as a noun) and in favour of BGH (which accords the gerund the status of a noun formed by adding the inflectional ending -*Wa/no* to a verb stem).

4.5.2 Argument two

A -*Wa/no* word functioning as a participle cannot modify a DET C NUM′ sequence, C being the zero noun introduced in chapter 3 that carries the features count and inanimate; cf. (36-i). Placing the N after the NUM′ (which, in the case of (36-i), also means refraining from DENOM deletion) slightly improves the status of the expression; cf. (36-ii). Varying the DENOM has no effect (we save space by not giving examples of this).

(36) i. *apnar deWa je C Ta
"your given which Ø item"
'the one given by you which'
ii. ?apnar deWa je Ek Ta C
"your given which one item C"
'the one given by you which'
iii. 'the two given by you which': a. *apnar deWa je C du To
b. ?apnar deWa je du To C
iv. 'the three given by you which': a. *apnar deWa je C tin Te
b. ?apnar deWa je tin Te C

Of course, the (a)-(b) differences do not appear in elicitable judgements of grammaticality. The * judgement for the (a) cases gets overridden by the ? judgement for the (b) cases, as in any choice between a better and a worse judgement for a single terminal string with two different surface structures. So we extrapolate from the (i)-(ii) difference which we do observe to the (a)-(b) difference which we cannot. Recall that the *Ta* in (i) is the output of the rule of *Ek*-Deletion, which reduces [$_{NUM'}$ [$_{NUML}$ Ek][$_{DENOM}$ Ta]] to [$_{NUM'}$[$_{DENOM}$ Ta]] after a noun.

The fact to be accounted for, then, is that a participle ending in -*Wa/no* cannot modify a DET C NUM′ sequence. The judgement about DET NUM′ C is less clear and therefore best left for later research. How well can BGH and TH handle the fact about DET C NUM′?

BGH assigns to (36-i), (36-iii-a), and (36-iv-a) base-generated and unaltered structures of the form exemplified in (37-i). TH assigns instead surface structures like (37-ii-a) derived from deep structures like (37-ii-b).

(37) i. [$_{NP}$ [$_{AP}$ [$_{NP}$ apnar][$_{A}$ [$_{N}$ deWa] Ø]][$_{DET}$ je][$_{NUM'}$ [$_{NUML}$ du][$_{DENOM}$ Ta]]]
ii. a. [$_{NP}$ [$_{S}$ [$_{NP}$ apnar][$_{V}$ deWa]][$_{NP}$ [$_{DET}$ je][$_{N}$ C[$_{NUM'}$ [$_{NUML}$ du][$_{DENOM}$ Ta]]]]
b. [$_{NP}$ [$_{S}$ [$_{NP}$ apnar][$_{NP}$ [$_{DET}$ je][$_{N}$ C[$_{NUM'}$ [$_{NUML}$ du][$_{DENOM}$ Ta]]]
[$_{V}$ deWa][$_{NP}$ [$_{DET}$ je][$_{N}$ C[$_{NUM'}$ [$_{NUML}$ du][$_{DENOM}$ Ta]]]]

For the purposes of this part of the argumentation, the NUM′ Postposing analysis of the origin of N NUM′ sequences plays no role; for the moment, I shall ignore the question of where these sequences come from; the answer makes no difference in this specific context.

BGH, yielding as it does the surface structure (37-i), makes it possible to account for the data without postulating any special mechanism marking (37-i) as ill-formed. [$_{NP}$ AP DET C NUM′] as such happens to be ungrammatical regardless of whether the AP is built around a participle. This more general fact may reflect strict subcategorization, or a filter, or some other principle. Be that as it may, BGH can subsume the case at hand, *Participle Phrase DET C NUM′, under the general prohibition, *AP DET C NUM′, and thus need not posit any mechanism peculiar to participles. A non-participle example of *AP DET C NUM′ is **Sobuj e C Ta* "green this C item" 'this green one'.

But TH, since it postulates surface structure (37-ii-a) (derived from (37-ii-b), does not allow one to subsume the fact that *apnar deWa je C du To* is ill-formed under a broader fact.

This is not obvious at once. One might think that a proponent of TH could reanalyze *Sobuj e C Ta* as an [$_{NP}$ [$_{S}$ A][$_{NP}$ DET C NUM′]] structure and then generalize the

ungrammaticality of **apnar deWa je C du To* and **Sobuj e C Ta* to that of [$_{NP}$ S [$_{NP}$ DET C NUM′]]. But such a strategy would fail. An analysis of adjectives as embedded clauses would, to be consistent, carry with it an [$_{NP}$ S NP] analysis of relative clauses (at this stage of the argument, the S vs S′ distinction, or the COMP node, does not matter). Thus, a proponent of the [$_{NP}$ [$_{S}$ A] DET C NUM′] analysis of **Sobuj e C Ta* would be committed to an [$_{NP}$ S NP] analysis of, say, [$_{NP}$ [$_{S}$ [$_{NP}$ [$_{DET}$ je][$_{N}$ kaThi][$_{NUM′}$ Ta]][$_{A}$ Sobuj][$_{V}$ chilo]][$_{NP}$ [$_{DET}$ Se][$_{N}$ C][$_{NUM′}$ Ta]]] "which stick item green was that C item" 'the stick which was green'. But *je kaThiTa Sobuj chilo Se C Ta* is well-formed. Therefore, [$_{NP}$ S [$_{NP}$ DET C NUM′]] as such cannot be prohibited. Since (37-iii) and (37-iv), which one wishes to prohibit, share no feature which such a structure for *je kaThiTa Sobuj chilo Se C Ta* lacks, a proponent of TH thus cannot postulate a general prohibition covering both **Sobuj e C Ta* and **apnar deWa je C du To.*

In other words, TH misses the generalization and must set up special machinery to rule out (37-iii) and, separately, (37-iv). Clearly, BGH does a better job.

(37). iii. *[$_{NP}$ [$_{S}$ [$_{A}$ Sobuj]][$_{NP}$ [$_{DET}$ e][$_{N}$ C][$_{NUM′}$ Ta]]]
'this green one'
iv. *[$_{NP}$ [$_{S}$ [$_{NP}$ apnar][$_{V}$ deWa]][$_{NP}$ [$_{DET}$ je][$_{N}$ C][$_{NUM′}$ [$_{NUML}$ du][$_{DENOM}$ Ta]]]]
'the two given by you which'

4.5.3 Argument three

Not all gerunds can also function as participles. Some cannot:

(38) i. ramer baje kOtha *bole bERano*
"Ram's nonsensical things *say-and going-around*"
'Ram's going around talking nonsense'
ii. *ramer *bole bERano* baje kOtha
"Ram's *say-and gone-around* nonsense"
'nonsense Ram has been talking'
iii. ramer Sastriyo Songgit *ERano*
"Ram's classical music *avoiding*"
'Ram's avoiding classical music'
iv. *ramer *ERano* Sastriyo Songgit
"Ram's *avoided* classical music"
'classical music avoided by Ram'

If one substitutes *phEla* 'dropped, dropping' for *bERano*, or *Sekha* 'learning, learnt' for *ERano*, then participle constructions as well as gerund constructions become admissible. But certain words resist the participle construction. These facts must be docketed as idiosyncratic information in the lexical entries of the relevant words.

One can draw two conclusions from the existence of gerunds unable to play the participle role.

First, it is better to record the gerund-participle relation in the lexicon, as BGH does, than to write it as a syntactic transformation, as TH does. A syntactic transformation *prima facie* expresses a general mapping which brooks no exceptions. In order to mitigate the effect

of the transformation, TH needs to add to it a series of afterthoughts that exempt constructions containing specific lexical items from undergoing the transformation. BGH needs no afterthoughts. BGH states the gerund-participle relation as a lexical rule in the first place. So, each gerund's lexical entry is marked as to whether it undergoes the gerund-participle rule. This is appropriate. The job of a lexical rule is to express a subregularity, a partial mapping. Items that are party to such a mapping have to say so in their lexical entries. And we have just seen that only some gerunds also act as participles. So, whatever one's theory, one must make one's lexical entry for each gerund declare whether it is or is not one of those dual-role gerunds. If one happens to be a proponent of BGH, one will couch this declaration in terms of the applicability or inapplicability of the gerund-participle rule, a lexical rule whose functioning requires such declarations. If one believes in TH instead, one will need to state whether or not particular gerund words are exceptions to the Equi-NP transformation – a rule which is not expected to permit any exceptions. Thus, the facts again strengthen BGH and weaken TH.

The second conclusion is entailed by the existence of gerunds which do not permit participle use in conjunction with the non-existence of any *-Wa/no* words which are used as participles but not as gerunds. The conclusion is this. Rule (32), which says that some gerunds, marked as such, also act as participles, is the right generalization. It would be wrong to turn it around and say that some participles, marked as such, also act as gerunds. It would also be wrong to say that there is a long list of gerunds and a long list of participles and these lists happen to overlap in part. I will first explain why these statements would be wrong, one by one. Then I will put the matter more technically.

Suppose you say that there are two sets of entries, among others, in the Bangla lexicon: P, the set of participle entries, and G, the set of gerund entries. And you note that some overlap exists. You formulate a rule which says that if P_i, a member of P, and G_i, a member of G, happen to be the same word, then the 'subject' of P_i must match the 'direct object' of G_i, i.e. the class of 'subjects' that P_i can have must be the same as the class of 'direct objects' that G_i can have. I will call this proposal the Overlap Analysis, OA.

The problem with OA is that it is too weak. It is consistent with the facts of Bangla, but it would be compatible also with the facts of German, if we were to take the liberty of calling forms in *-ung* (e.g. *Übersetzung* 'translation') 'gerunds'. German has a set of 'gerunds', then, and a set of past participles which are indeed traditionally called past participles. OA predicts that if any German 'gerund' and any German past participle happen to coincide then certain entailments will hold. The trouble is, the intersection set is empty: there is no German word which belongs both to G and to P. However, the fact that this intersection is empty does not refute OA. No German construction has any property that OA says it should not have. Since we want an account of gerunds and participles in Bangla not to be so consistent with the facts of the radically different language German, we must reject OA precisely because it is impossible to refute it in German.

Suppose you now try an analysis which does distinguish Bangla from German. You allow Bangla just one set of lexical entries, L, where German has two, G and P. You arbitrarily choose to put the participle function in the lexical entries and to arrive at the gerund function through a lexical rule. All members of your L, then, are Adjectives of the form V + *Wa/no*. A lexical rule – not a gerund-participle rule, but a participle-gerund rule – forms gerunds from such participles and switches subjects into direct objects. Your proposal, which may be called the Participle-Gerund Analysis or PGA, fares better than OA did. Unlike OA, PGA guarantees

a non-empty set each of whose members will be both a participle and a gerund.

PGA, however, fails to generate those gerunds, like *ERa-no* 'avoid-ing', which are always gerunds and therefore cannot be derived from participles (cf. **ramer ERano Sastriyo Songgit* 'classical music avoided by Ram'). This observational inadequacy already suffices to condemn PGA. But it may be worth mentioning another defect. PGA has a lexical rule forming gerunds from participles. Each participle must choose whether or not to undergo the rule which, being a lexical rule, does not automatically apply to all participles. For the choice to be real, there must be some participles which choose not to undergo the rule. Thus, PGA predicts the existence of participles that do not also act as gerunds. This prediction is false. Notice that this false prediction is also made by OA.

The rejection of OA and PGA may be summed up as follows in ordinary language. The facts are that some gerunds are just gerunds, others are also participles, and every participle is also a gerund. PGA misses the gerunds that are just gerunds. OA, like the existentialists' nothingness, permits everything and specifies nothing. Only rule (32), coupled with a lexicon that registers gerunds but not participles, gets the facts straight:

(32) $\overset{\text{Ø}}{\ }[[_{\text{N}}\,\text{V} + \text{Wa/no}]]_{\text{A}}\ (\text{II}) = \text{I}$

The lexical rule needed for PGA would be (32′), the reverse of (32), in Wasow's (1977) notation. The equivalence rule needed for OA could be written as (32″), where the parenthesis-comma notation indicates ordered pairs; (32″) is my attempt to say that the relation of a II to its gerund equals that of a I to its participle.

(32′) $\overset{\text{Ø}}{\ }[[_{\text{N}}\,\text{V} + \text{Wa/no}]]_{\text{A}}\ (\text{I}) = \text{II}$

(32″) $(\text{II}, [_{\text{N}}\,\text{V} + \text{Wa/no}]) = (\text{I}, [_{\text{A}}\,\text{V} + \text{Wa/no}])$

Wasow (1977), whose notation I have used in (32) and (32′), had no way to select the correct 'direction' for the lexical rule he formulated in English. He could not, in fact, be sure if the question of direction was an empirical one. But the Bangla data show that the non-directional analysis (32″) is unacceptable and that, of the directional analyses (32) and (32′), (32) is better than (32′). So, the question of direction for such lexical rules is empirical. With luck, it can even be answered in some cases.

4.5.4 Argument four

So far, we have only looked at cases where the direct object of a transitive gerund corresponds to the head nominal modified by the corresponding participle, the participle belonging to the type traditionally called Past Passive. Consider now cases where the subject of a verb, instead of its direct object, is involved.

(39) i. dilli theke phire aSa tin jon montri
"Delhi from back come three fella minister"
'three ministers who have come back from Delhi'

ii. tin jon montrir dilli theke phire aSa
"three fella minister-GEN Delhi from back coming"

‘the return of three ministers from Delhi’

iii. ingreji jana maSTar
“English knowing teacher”
‘a teacher who knows English’

iv. maSTarer ingreji jana
“teacher’s English knowing”
‘the / a teacher’s knowing English’

Both in (30-i), where the verb stem *aS* is transitive, and in (iii), where the verb stem *jan* is intransitive, the participle can be characterized as ‘active’. It is difficult to provide uncontroversial specifications of tense. One might feel inclined to suggest that *aSa* is a past participle and *jana* a present participle, but such decisions would be hard to reconcile with (v):

(39) v. ingreji Sekha chatro
“English learnt student”
‘a student who has learnt English’

vi. chatror ingreji Sekha
“student’s English learning”
‘the student’s learning English’

Since it would appear that the appearance of *Sekha* being a ‘past’ participle merely results from the difference in meaning between *Sekh* ‘learn’ and *jan* ‘know’, I hesitate to claim that active participles in Bangla are specified for tense over and above what the meaning of a verb stem imposes on its participle.

Active participles are rare. The extent of the construction remains unresearched. We do not know what subclass of gerunds can also be active participles. One can easily show that some gerunds cannot (cf. (vii)) and that a gerund may act both as a passive and as an active participle (cf. (ix) below and (iv) above).

(39) vii. *dowRono maSTar
“running/run teacher”
‘a teacher who has run or is running’

viii. maSTarer dowRono
“teacher’s running”
‘the teacher’s running’

ix. maSTarer jana ingreji
“teacher’s known English”
‘the English that the teacher knows’

In the absence of any insight about exactly which gerunds can also be active participles, I take it that the gerunds that do so carry a lexical marking to that effect and that the marking subjects such gerunds to rule (40), a lexical rule distinct from (32).

(40) Gerund / Active Participle Rule

$$\text{Ø}_{[[_{N}\ V + \text{Wa/no}\ ]]_{A}}\ (\text{I}) = \text{I}$$

This natural extension of BGH handles the phenomena shown in (39) in a plausible fashion. The Transformational Hypothesis yields no natural extension which could adequately handle the data.

4.5.5 Argument five

Yet another cognate data set which covers only part of a lexical domain:

(41) i. amar toyri baRi
"my built house"
'a house built by me'
ii. amar baRi toyri
"my house building"
'my building a house'
iii. Omiter ranna bhat
"Amit's cooked rice"
'rice cooked by Amit'
iv. Omiter bhat ranna
"Amit's rice cooking"
'Amit's cooking rice'
v. nOSTo SomOY
"wasted time"
'time wasted'
vi. SomOY nOSTo
"time wast(ing)"
'wast(ing) of time'
vii. poriSkar bonduk
"cleaned gun"
'a cleaned gun'
viii. bonduk poriSkar
"gun cleaning"
'cleaning of a gun'
ix. bharoter amdani ca
"India's imported tea"
'tea imported by India'
x. bharoter ca amdani
"India's tea import"
'India's import of tea'

Again, there are exceptions, such as (xi-xii). The scope of the pattern is unknown.

xi. *peS prostab
"presented proposal"
xii. prostab peS
"proposal presenting"

'presenting a proposal'

The forms of (41) are related to conjunct forms like *amar toyri kOra baRi* "my building done house" 'a house built by me' and *amar baRi toyri kOra* "my house building doing" 'my building a house'. But the correspondence is inexact. Ill-formed (41-xi) corresponds to well-formed *peS kOra prostab* "presenting done proposal" 'a proposal presented'. So, an analysis of the conjunct forms, even if available, would not solve the problems of (41). To reinforce this point, consider the pair *protimaY debir odhiSThan* "image-LOC goddess-GEN presence" 'the presence of the goddess in the image' and *protimaY odhiSThito debi* "image-LOC present goddess" 'the goddess present in the image' – here the noun *odhiSThan* and the adjective *odhiSThito* differ somewhat in form, but the pair has the same properties as the pairs in (41). However, this pair has no conjunct paraphrase. I therefore set aside the forms with conjunct verbs in them.

Is it BGH or is it TH that is better equipped to produce a natural extension to handle (41)?

BGH, only slightly modified, does the job. One need only simplify (32) to (42) and, apart from the markings on many gerunds making them subject to the rule, one must now also mark a few more nouns like *toyri* and *ranna* (but not *peS*) as subject to the rule now formulated as (42). By contrast, TH would need to be complicated, not simplified, in order to accommodate (41).

(42) Simplified Gerund/ Passive Participle Rule

$$\emptyset_{N_A}\ (II) = I$$

4.6 On the verbal character of gerunds

Section 4.5 has shown the superiority of BGH to TH. Recall that the two differed mainly in that TH accorded gerunds the status of just verbs while BGH regarded gerunds as nouns derived by adding to a verb stem the inflectional ending *-Wa/no*. Insofar as the hypotheses differ, section 4.5 has shown that BGH is correct. Insofar as they agree – and they do agree that gerunds are at least verbs if not (as BGH, contrary to TH, adds) more – they both predict that a gerund should share properties of finite verbs not shared by non-gerund nouns. The facts, of course, bear out this prediction made by both hypotheses. One may think that we need not cite the facts; they cannot advance the present inquiry. But some of the facts deserve to be on record, I feel, because they undermine the widespread understanding of 'properties of verbs not shared by non-gerund nouns'.

The subclass of nouns exemplified in (41) serve as a useful control group. I will call them Quasi-gerunds. They display what one might see as some of the more glaring properties of verbs. They can take (i) direct objects as in (41) and as in (43-i), where the object even carries the Case marker *-ke;* (ii) indirect and direct objects as in (43-ii); (iii) adverbs as in (43-iii) and spatio-temporal adjunct PPs as in (43-iv); (iv) and even subordinate clauses of the form $[_{PP}$ S P] as in (43-v). Cf. Chomsky's (1970: 193) point about how *criticism* in English can never take a *before*-clause although *criticizing* can.

(43) i. rajake Opoman
"king-OBJ insult"
'an insult to the king'

ii. biSSobiddalOYke pOncaS hajar monjur
"university-OBJ fifty thousand grant"
'granting the university 50000 (sc. rupees)'

iii. SOmoSSar cOTpOT mimangSa
"problem's quickly solution"
'a quick resolution of the problem'

iv. adh ghOnTar moddhe mangSo ranna
"half hour's within meat cooking"
'cooking meat within half an hour'

v. [$_{PP}$ [$_{S}$ abduler SOngge rajnoytik mOte mele na][$_{P}$ bole]] tar kobitar SOmalocona
"Abdul-GEN with political opinion-LOC agrees not because his poetry's criticism"
'criticizing Abdul's poetry because one doesn't agree with him politically'

However, the subclass of nouns – quasi-gerunds – which have these properties are not verbs in any sense; quasi-gerunds are not gerunds. That they have the properties exemplified above goes to show that these properties do not work as diagnostics of verbhood in Bangla.

True diagnostics of verbhood include: (i) negatability, (ii) taking a nominative subject, (iii) taking a vector (forming a compound verb). Gerunds, like finite verbs, pass these tests, which quasi-gerunds fail.

Negatability. Verbs, and verbs alone, can be negated with a single *na*.

(44) i. amar caYe du camocer beSi cini lage na
"my tea-Loc two spoon-Gen more sugar will-be-needed not"
'I won't need more than two spoonfuls of sugar in my tea'

ii. 'I will need not more than two spoonfuls of sugar in my tea':
*amar caYe du camocer beSi na cini lagbe
*amar caYe du camocer na beSi cini lagbe
*amar caYe na du camocer beSi cini lagbe

Gerunds can be *na*-negated; cf. (iii). Quasigerunds cannot; cf. (iv):

iii. Omiter bhat na raMdha
"Amit-Gen rice not cooking"
'Amit's not cooking rice'

iv. *Omiter bhat na ranna
"Amit-GEN rice not cooking"

Nominative subject. Verbs, only verbs, may take a nominative subject.

(45) i. ami jabo
"I will-go"

'I will go'

ii. *[ami jatra]
"I journey"

Gerunds may take a nominative subject and sometimes must; cf. (iii-viii). Quasigerunds never take a nominative subject; cf. (ix).

iii. ami jaWar ceYe tumi jaWa-i bhalo
"I going-GEN than you going-EMP good"
'It is better for you to go than for me to go'

iv. amar jaWar ceYe tomar jaWa-i bhalo
"my going-GEN than your going-EMP good"
'It is better for you to go than for me to go'

v. tumi amake janiYe thaka SOtteW
"you me informed having despite"
'despite your having informed me'

vi. ?tomar amake janiYe thaka SOtteW
"your me informed having despite"
'despite your having informed me'

vii. gOromkale SaRe chOTaY rod na thaka haSSokOr
"summer-Loc half-past six-Loc sunlight not being ridiculous"
'It is ridiculous for there to be no sunlight at 6.30 in summer'

viii. *gOromkale SaRe chOTaY roder na thaka haSSOkOr
"summer-LOC half-past six-LOC sunlight-GEN not being ridiculous"

ix. *Omit bhat ranna
"Amit rice cooking"

Vector. Verbs, only verbs, may take a vector:

(46) i. eSa khOborTa janabe
"Esha news-the will-report" (no vector)
'Esha will report the news'

ii. eSa khOborTa janiye debe
"Esha news-the report-and will-give" (vector present)
'Esha will report the news "perfectively"'

Gerunds may take a vector, cf. (iii)-(iv), and sometimes – in the case of stems which in the non-nominal conjugation require a vector – must, cf. (v)-(vi). Quasi-gerunds never take a vector. It is difficult to exemplify the ill-formedness of a non-verb taking a vector, because one cannot imagine a conjunctive for a non-verb; (viii) is an attempt to imagine one.

iii. Omiter bhat raMdha
"Amit's rice cooking"
'Amit's cooking rice'

iv. Omiter bhat reMdhe deWa
"Amit's rice cook-and giving"

‘Amit’s cooking rice for (someone)”

v. *kukur lElano
“dog sicking”

vi. kukur leliye deWa
“dog sick-and giving”
‘sicking dogs (on someone)’

vii. Omiter bhat ranna
“Amit’s rice cooking”
‘Amit’s cooking rice’

viii. *Omiter bhat ranne deWa
(from **/ranna-ye/, double-starred because the question of
attaching a conjunctive suffix to a non-verb simply does not arise)

So, the fact that gerunds, unlike quasi-gerunds, contain an internal V node makes for systematic differences in syntactic behaviour. Some of these differences seem to call for a slight modification of our hypothesis that gerunds are nouns of the form V+*Wa/no*. It appears that we should have V′ there instead of V. This modification accommodates *na*-negation, vectors, and a combination of the two as in *Omit bhat reMdhe na deWa SOtteW* “Amit rice cook-and not giving despite” ‘despite Amit’s not cooking rice for (someone)’.

I have no explanation for the nominative subject phenomena. In this study I have taken no position on how Bangla interrelates Case forms with grammatical functions. Miriam H. Klaiman of the University of Chicago has informed me that she is investigating this question thoroughly. Appropriately enough, her work is diachronic. Case phenomena in Bangla exhibit idiosyncrasies due to uneven patterns of change.

4.7 Back to infinitives

Having surveyed, in sections 4.3-4.6, the properties of Bangla gerund forms and their part-time participle function, I now return to the relation between the gerund and the infinitive. Section 4.2 suggested, on the basis of (17)-(20), that the gerund blocks the infinitive in subject position and is blocked by the infinitive in complement position. Armed with our new awareness of subject-laden gerund constructions, we will now look at successively stronger groups of counterexamples to the initial guess based on (17)-(20). The counterexamples refute the idea that infinitives block gerunds in complement position – an idea which constituted half of the conjecture in section 4.2.

(47) i. amra gaRi calano Sikhchi
“we car driving are learning”
‘We are learning how to drive a car’

ii. amra gaRi calate Sikhchi
“we car to-drive are-learning”
‘We are learning how to drive a car’

iii. rupen SaMtar Sekha Suru koreche
“Rupen swimming learning commencement has-done”
‘Rupen has started learning how to swim’

iv. rupen SaMtar Sikhte Suru koreche
"Rupen swimming to-learn commencement has-done"
'Rupen has started to learn how to swim'

v. rupen bichana kOra SeS koreche
"Rupen bed-making end has done"
'Rupen has finished making his bed'

vi. *rupen bichana korte SeS koreche
"Rupen bed to-make end has-done"
'*Rupen has finished to make his bed'

vii. amra ramer jaWa caybo na
"we Ram's going will-want not"
'We will not want Ram to go'

viii. *amra ramer jete caybo na
"we Ram's to-go will-want not"

The guess was that in complement position infinitives blocked gerunds. But (i)-(ii) show that in at least one case both gerund and infinitive forms may occur in complement position; neither form blocks the other. This substitutability, (iii)-(iv) show, is not confined to one pair of examples. In (v)-(vi) not only does the infinitive fail to block the gerund, but the gerund appears to block the infinitive, in examples chosen to illustrate a similar phenomenon in the grammar of the English glosses. Finally, (vii)-(viii) again give the impression that the gerund is blocking the infinitive rather than the other way round. Note that the gerund construction in (vii) has a subject. All subject-laden gerunds behave as *ramer jaWa* in (vii) does: they all have the property that the infinitive cannot replace them in complement position.

With increasing force, the examples refute our first guess about the relation between complement gerunds and complement infinitives. The examples also suggest that no single proposal will adequately replace the initial conjecture. Examples (i)-(vi) show that, at least in part, the matrix verb determines the possibility of using a gerund, an infinitive, or both. But (vii)-(viii) point up the impossibility of relegating the whole problem to the lexicon, since the general fact that the complement position tolerates subject-laden gerunds but never subject-laden infinitives, being general, does not belong in the lexicon. We set aside the lexical properties of matrix verbs and concentrate on the general fact about subject-laden gerunds apparently blocking subject-laden infinitives.

A quick test shows that subject-laden infinitives do no better in subject position than in complement position; cf. (47-ix, x). Thus, the overall distribution is as in (48).

(47) ix. ramer jaWa SOmbhOb hObe na
"Ram's going possible will-be not"
'For Ram to go will not be possible'

x. *ramer jete SOmbhOb hObe na
"Ram's to-go possible will-be not"

(48)

	GERUNDS		INFINITIVES	
	SUBJECTLESS	SUBJECT-LADEN	SUBJECTLESS	SUBJECT-LADEN
Subject	occur	occur	missing	missing
Complement	(occur)	occur	(occur)	missing

We have agreed to relegate to the lexicon the fact that subjectless gerunds and/or subjectless infinitives in complement position get chosen by the matrix verb. Hence the parentheses in (48), indicating that the parenthesized part of the pattern is handled without appeal to the concept of blocking. The rest of the pattern, though, can be characterized by saying (i) that, in both subject and complement position, subject-laden infinitive phrases are blocked by subject-laden gerund phrases and (ii) that, in subject position, subjectless infinitive phrases are blocked by subjectless gerund phrases. These are both general statements. But the formulation I have given is neutral between several ways of incorporating the statements into a grammar as formal generalizations. We want to choose between some possible formalizations of the properties of infinitives unearthed in this section. To do this we need to return to the question of the place of infinitive structures in modal constructions.

4.8 The syntactic structure of the modal construction

Chapter 3 established that the syntactic structure of the modal construction differed from that of the compound verb, but left open the question of what the structure of the modal construction was. We will now deal with this question. Without some answer to it we cannot formalize the general statements (i) that subject-laden gerund phrases block subject-laden infinitive phrases in both subject and complement position and (ii) that subjectless gerund phrases block subjectless infinitive phrases in subject position.

The unit movement test for constituent structure tells us two things which seem to contradict each other. One preposing rule relates (i) and (ii) below, suggesting that the sequence of infinitive verb and modal forms a constituent. Another preposing rule relates (iii) to (i), suggesting that the sequence of NP and infinitive forms a constituent. Thus, the unit movement test gives evidence in favour of both the structures (iv) and (v) for sentence (i).

(49) i. ke oke e-SOb ante bollo?
"who? him/her this-all to-bring told"
'Who told him/her to bring all this?'

ii. *ante bollo* ke oke e-SOb?
"to-bring told who? him/her this-all"
'Who *told* him/her *to-bring* all this?'

iii. *e-SOb ante* ke oke bollo?
"this-all to-bring who? him/her told"
'Who told him/her *to bring all this*?'

iv. ke oke e-SOb [[ante][bollo]]
"who? him/her this-all [[to-bring] [told]]"

v. ke oke [[e-SOb][ante]] bollo
"who? him/her [[this-all] [to-bring]] told"

All linguists who use movement rules agree that the unit movement test is the best diagnostic we have for constituency. So, let us build an analysis which enables both (iv) and (v) to be the structure – at different stages in a derivation, of course – of (49-i). The pivot of such an analysis is a transformation interrelating (iv) and (v). I propose that this transformation

operates on (vi) to yield (vii), by adjunction. (50) formalizes the rule involved.

(49) vi. $[_S\,[_{NP}$ ke$][_{NP}$ oke$][_{VP}\,[_{NP}$ e-SOb$][_V$ ante$]][_{V'}[_V$bol$][_T$ l$][_E$ o$]]]$
vii. $[_S\,[_{NP}$ ke$][_{NP}$ oke$][_{NP}$ e-SOb$][_{V'}[_V$ ante$][_{V'}[_V$bol$][_T$ l$][_E$ o$]]]]$

(50) i. S → NP NP VP V′
ii. VP → NP V
iii. V′ → V T E
iv. Infinitive reanalysis (optional)
S.D.: X – V – V′
S.C.: 1 2 3 ➔ 1, 0. 2+3

Before trying to find support for this analysis, I need to note that Ian Smith (personal communication) has proposed that (49-ii) should be derived from (49-i) not by preposing of *ante bollo* but by one-by-one postposing of *ke*, *oke*, and *e-SOb*. To reproduce his statement more accurately, he has pointed out to me that in Srilankan Tamil and other Srilankan languages one can motivate a proposal to this effect. The situation in those Srilankan languages differs from that in Bangla in crucial ways. As far as I can judge, there is no support for a proposal like Smith's in Bangla. But one cannot legislate against someone making a proposal and promising to motivate it at some later date. To readers who want to take this course I must point out that independent data, pertaining to Gapping, leads to the conclusion which I have based on the idea that (49-ii) involves preposing. Thus, the conclusion is a firm one. (50) is *prima facie* well-motivated.

Now for further support. The transformation as it stands correctly derives (51-i). The only part of the analysis which needs to stretch to accommodate the complexity of (51-i) is rule (50-ii), which I restate as (50-v):

(50) v. VP → (NP) (VP) V

(51) i. $[_S\,[_{NP}$ puliS$][_{VP}\,[_{VP}\,[_{NP}$ corke$][_{VP}\,[_V$ palate$]][_V$ dite$]][_V$ cayte$]][_{V'}\,[_V$ par$][_T$ Ø$][_E$ e$]]]$
"police thief-OBJ escape let want may"
'The police may want to let the thief get away'
ii. $[_S\,[_{NP}$ puliS$][_{VP}\,[_{NP}$ corke$][_{VP}\,[_V$ palate$]][_V$ dite$]][_{V'}\,[$cayte$][_{V'}$ pare$]]]$
iii. $[_S\,[_{NP}$ puliS$][_{NP}$ corke$][_{VP}\,[_V$ palate$]][_{V'}\,[_V$ dite$][_{V'}[_V$ cayte$][_{V'}$ pare$]]]]$
iv. $[_S\,[_{NP}$ puliS$][_{NP}$ corke$][_{V'}\,[_V$ palate$][_{V'}\,[_V$ dite$][_{V'}[_V$ cayte$][_{V'}$ pare$]]]]$

Let me go over derivation (51) in some detail. A mechanical problem arises. In view of the fact that *cayte* 'want' and *pare* 'may' in (ii) match terms 2 and 3 in the S.D., can't the rule adjoin *cayte* to *pare* when it applies to (ii)? If it does so, what happens? If we adopt Bresnan's (1976a) Relativized A-over-A Condition then the rule is indeed entitled, in (ii), to adjoin *cayte* to *pare* instead of adjoining *dite* to *cayte pare*. The immediate result of such an application will be (51-y).

(51) y. $[_S\,[_{NP}$ puliS$][_{VP}\,[_{NP}$ corke$][_{VP}$ palate$]][_V$ dite$]][_{V'_1}\,e\,[_{V'_2}\,[_V$ cayte$][_{V'_3}$ pare$]]]]$

Evidence will be presented later that at least the rule of Infinitive Reanalysis does not

leave trace, but removes the V node and its contents, without residue, from the node's original position. Thus the old V'_1 is left exhaustively dominating the new V'_2 node conjured up by the adjunction process. Traditional practice calls for 'pruning' of V'_1, yielding (51-z), which, of course, is identical to (51-ii).

(51) z. $[_S\,[_{NP}\,puliS][_{VP}\,[_{NP}\,corke][_{VP}\,[_V\,palate]][_V\,dite]][_{V'_2}\,[_V\,cayte][_{V'_3}\,pare]]]$

Thus, the application of rule (50-iv) to *cayte* in (51-ii) makes no dent in the structure of this phrase-marker.

At this point one may demand clarification of or justification for the traditional pruning convention. Traditional pruning and present-day adjunction (which Ross (1967) called 'Chomsky-adjunction' and Baker (1978) calls 'aunt-adjunction') are two sides of the same coin. Pruning requires that a node A which, at some point in the derivation, has only an A daughter be deleted, as in (52-P). Adjunction requires that a node A and a non-A node B adjoined to A receive a new mother A: (52-A).

(52-P) Pruning: (p) (q)

(52-A) Adjunction: (p) (q)

How exact the correspondence between (52-P) and (52-A) is becomes clear if one explicitly shows the everywhere-present concatenational element *e* in B's position. (52-P) then appears as (53-P).

(53-P) Pruning: (p) (q)

Comparing (52-P) and (53-P) one sees that in principle (p) and (q) are equivalent in both cases, but that as a matter of notation one must write the A_1 if A_2's sister is non-null whereas one must omit the A_1 if A_2's sister is null. Pruning – the convention which chooses notation (q) in (53-P) – and adjunction – the convention which chooses notation (p) in (52-A) – both express the notational principle that A_1 should be written (notation (p) should be used) if and only if A_2 has at least one non-null sister.

I take it to be uncontroversial that the current procedure of adjunction is correct. Baltin

(1978) gives empirical arguments in its favour. Therefore, this procedure's alter ego, the traditional pruning convention which maps derivations into phrase-markers, is correct by the same token.

Here ends the digression about why applying rule (50-iv) to a non-maximal V′ has no ill effects (no discernible effect at all, in fact). Resuming our discussion of (51), we note that the pruning of the top VP on the way from (i) to (ii), of the middle VP on the way to (iii), and of the lowest (innermost) VP on the way to (iv) does not exemplify the traditional pruning convention discussed above. Rather, derivation (51) makes assumptions (54-i) and (54-ii), both unrelated to the traditional pruning convention. (54-ii) is justified in Schwartz (1972).

(54) i. The rule of Infinitive Reanalysis removes the V node and its contents, without trace, from its position as head of VP.
ii. When a VP loses its head V, the VP node is 'pruned'.

These assumptions need justification. If VP Deletion is responsible for the gap after *wasn't* in *Hector was sober but Helen wasn't*, then *sober* in the first conjunct is a VP, even though it has lost its V *be* to the AUX node. And Selkirk's (1972) explanation for the difference between *He'll never see it but she will* and **He'll never see it but she'll* casts doubt on (54-i). Without taking any position on the validity of these proposals for English, I will submit evidence that the surface structures which (54) predicts for Bangla are well-founded and that therefore so is (54). Whether this conclusion implies a need to re-examine current proposals about the English verb phrase is a question that I will leave open.

If one rejects a theory that incorporates (54), one may embrace any of several conceivable alternatives, of which three seem to merit consideration.

Alternative X. (54-i) is false. Infinitive Reanalysis leaves *e* to mark the original site of the V. But (54-ii) is true. If a VP has no tangible head, the VP node is deleted.

Alternative Y. (54-i) is false. Infinitive Reanalysis leaves trace as in X. And (54-ii), though true in the sense that a VP which dominates no V node would be deleted, is irrelevant, since the output of Infinitive Reanalysis still has a V under the VP. So, the VP node remains.

Alternative Z. (54-i) is true. Infinitive Reanalysis leaves no trace. But (54-ii) is false. The VP node, though it dominates no V, remains.

I have no factual argument to distinguish my view from alternative X. I can only point to the absence of any arguments for the putative null V in the relevant Bangla sentences, and appeal to simplicity. Luckily, X is a straw man unlikely to be maintained by any real person.

Y and Z, though they differ, agree that VP remains unpruned in the output of Infinitive Reanalysis. This shared claim of Y and Z is what I will argue against. If one agrees with my methodological predilection against X, then one will take the falsity of Y and Z to suggest that my analysis is correct.

Consider sentence (55-i). Views Y and Z, in their various ways, predict that VP Preposing will transform (i) into (ii). In contrast, my view predicts that, since *corke tala* is never a VP in (55-i), VP Preposing will fail to derive (55-ii). It is my prediction that is borne out by the facts.

(55) i. puliS $[_{VP}[_{VP}[_{NP}$ corke$][_{VP}[_{NP}$ tala$][_{V}$ bhangte$]][_{V}$ dite$]][_{V}$ cayte$]]$ pare
"police thief-OBJ lock break let want may"
'The police may want to let the thief break the lock'

ii. *[$_{VP}$ corke tala (..)] puliS bhangte dite cayte pare
"thief-OBJ lock police break let want may"
'The police may want to let *the thief* break *the lock*'

The (..) in (ii) represents the occurrences of null V which Y posits and Z does not posit. The VP bracketing shown in (i) is assigned by both Y and Z at the deep structure level.

To clinch the argument, I present yet another piece of evidence, foraying into the terrain of the analysis of Gapping in Bangla (which remains to be undertaken in earnest). (56) shows that gapping occurs only if the process leaves two (not more) constituents behind in the gapped structure. Thus, the failure of gapping in (57) shows that [corke tala] and [baRiWalake dOrja] are not constituents, and that, therefore, Y and Z, which lead to analyses treating these sequences as constituents, are wrong.

(56) i. input to gapping
ram SEmke cene ebong jodu modhuke cene
"Ram Shyam-OBJ knows and Jodu Modhu-OBJ knows"
'Ram knows Shyam and Jodu knows Modhu'
ii. output of gapping
ram SEmke ebong jodu modhuke cene
"Ram Shyam-OBJ and Jodu Modhu-OBJ knows'
'Ram knows Shyam and Jodu Modhu'
iii. input: ram SEmke Samanno cene ebong jodu modhuke moTamuTi cene
"Ram Shyam-OBJ little knows and Jodu Modhu-OBJ fairly-well knows"
'Ram doesn't know Shyam too well and Jodu knows Modhu fairly well'
iv. expected output turns out to be ill-formed: *ram SEmke Samanno ebong jodu modhuke moTamuTi cene
"Ram Shyam-OBJ little and Jodu Modhu-OBJ fairly-well knows"

(57) i. input: puliS corke tala bhangte dite cayte pare ebong bhaRaTe baRiWalake dOrja bhangte dite cayte pare
"police thief-OBJ lock break let want may and tenant landlord-OBJ break let want may"
'The police may want to let the thief break the lock and the tenant may want to let the landlord break the door'
ii. expected output turns out to be ill-formed: *puliS corke tala ebong bhaRaTe baRiWalake dOrja bhangte dite cayte pare
"police thief-OBJ lock and tenant landlord-OBJ door break let want may"
'*The police may want to let the thief break the lock and the tenant the landlord the door' (ill-formed in English)

The argument just presented rests on the assumption that the failure of gapping to apply to (57-i) does not have any reason distinct from the reason for the ungrammaticality of (56-iv). That assumption may prove shaky; in other words, it is possible that a careful analysis of gapping will invalidate this argument. Until such a turn of events stops us, however, it is reasonable to conclude that assumptions (54-i, ii) which derivation (51) relies on are indeed tenable and the obvious alternatives to (54) untenable.

As far as I can see, only one line of attack against derivation (51) remains unanswered. Namely, the careful reader may object that the /ye/-Insertion rule of chapter 3, which, given the S.D. V – V′, suffixes /ye/ to the first term, will apply to the output of Infinitive Reanalysis and perform illegitimate derivations like (58)-(59).

(58) [$_{V}$ cayte][$_{V'}$ pare]
"want may"
'may want'

(59) *[$_{V}$ [$_{V}$ cayte] /ye/][$_{V'}$ pare]

One way to meet this objection would be to order /ye/-Insertion before Infinitive Reanalysis. The output of the latter would then escape the action of the former. Phonologists call this sort of order 'counter-feeding order'. But that move would take us in the wrong direction. Not only is extrinsic ordering of transformations currently in disfavour; it would be too short-sighted a solution in this case. The real point is: how is it that a modal construction, although it ends up having the surface node structure of a compound verb, still refuses to behave the way compound verbs do? The reader may remember the fact – which my oversimple proposal (50) ignores – that the modal construction differs from the compound in that it tolerates double negation.

This objection makes me feel the need to subject the analysis offered in this section to a fairly thorough revision which, however, will leave the points made in this section intact. The syntactic structure of modal constructions has been ascertained. Section 4.9 will refine this analysis and solve the problem of (59) without compromising the results of section 4.8. Section 4.10 will return to the infinitive-gerund relation.

4.9 A detailed layout for the unaspected conjugation

This is going to be one of the few places in this volume where in order to get some work done I will proceed by unargued fiat rather than by weighing more than one option at a time. If in retrospect one is seen to have decided unwisely, one can recant and reinvestigate. For the sake of efficiency, I will also make no effort to point up the precise relation between the detailed proposals made in this section and the sketchy and somewhat different proposals made in earlier sections which this section supersedes. I will temporarily ignore negation.

(60) Unaspected non-nominal conjugation of /kOr/ 'do'

i. Untensed (non-finite) forms

Present Mood	Past Mood	Imperative Mood
korte	kore	korle

ii. Tensed (finite) forms: Simple Tense, Future Tense

ii-a. Present Mood

	1p.	2p.int.	2 p.n.	hon.p.	3 p.n.
Simple	kori	koriS	kOro	kOren	kOre
Future	korbo	korbi	korbe	korben	korbe

ii-b. Past Mood

Simple	korlam	korli	korle	korlen	korlo
Future	kortam	korti	korte	korten	korto

ii-c. Imperative Mood

Simple	kori	kOr	kOro	korun	koruk
Future	--	koriS	koro	korben	korbe

(61) Untensed structure: [$_V$V M], where M can be Present or Past or Imperative, respectively /yte ye yle/.

(62) Tensed structure: [$_V$V M T PERS (HON)], where M can be Present or Past or Imperative, T can be Simple or Future, PERS is whatever the paradigm dictates or allows, and HON (permitted only if PERS ≠ First Person) can be ± Honorific, [+Honorific] being actualized as /n/. [–Honorific] has distinctive manifestation only if PERS = Second Person, in which case it shows up as the second person intimate forms, distinct from the second person neutral forms which manifest the option of not having an HON node in the structure. In the Present and Past, particular M T sequences have invariant manifestations: Present Simple /Ø/, Present Future /yb/, Past Simple /yl/, Past Future /yt/. In the Imperative, each M T PERS sequence shows up as a fused form which in general resists decomposition.

The natural way to express the above, other things being equal, would be to write a PS rule V → V M (T) (PERS) (HON). Such a move, though, would precipitate chaos. The analyses collected in the present volume now show that Bangla correlates the presence of a T PERS (HON) sequence in a V exactly with that V being in a Main Verb position, i.e. __]$_S$. All and only sentential main verbs must have a T PERS (HON) sequence, then, after V M. As transformations are currently conceived, transformations do not serve our purpose, which is to specify this distributional fact. A new component is needed, an Inflection Component comprising Inflection Rules. This component follows the PS component but precedes lexical insertion and syntactic transformations. Inflection Rules are normally obligatory but may, at extra cost, be marked as optional. They apply in an extrinsically ordered sequence. For finite verbs the relevant Inflection Rules are (63)-(65). (In (65), the entire V from Y to Z counts as factor 2 in the Structural Description.)

(63) S. D. : [$_S$ X – V]
S. C. : 1 2 ➔ 1, 2 + [$_T$ ±Future]

(64) S. D. : [$_S$ X – V]
S. C. : 1 2 ➔ 1, 2 + [$_{PERS}$ ±Ego, ...]

(65) S. D. : [$_S$ X – [$_V$ Y [$_{PERS}$ –Ego, Z]]]

S. C. : 1 2 ➔ 1, 2 + [$_{HON}$ ± Honorific]

The Mood-introducing rules differ for finite and non-finite verbs. For finite verbs the rule is (66), ordered before (63). For the non-finite (specifically, infinitival) head of a VP the rule is (67). (68) specifies the Mood of the non-finite (conjunctive) forms in a compound verb. The relative ordering of (67), (68) and the finite verb rules makes no difference here. But one would probably invoke the normal cycle and have low domain rules apply before high-domain rules, where 'low' and 'high' refer to node height in phrase-markers.

(66) S. D. : [$_S$ X – V]

S. C. : 1 2 ➔ 1, 2 + $\left[_M \left\{ \begin{matrix} \text{Present} \\ \text{Past} \\ \text{Imperative} \end{matrix} \right\} \right]$

(67) S. D. : [$_{VP}$ X – V]
S. C. : 1 2 ➔ 1, 2+ [$_M$ Present]

(68) S. D. : [$_V$ V – V]
S. C. : 1 2 ➔ 1 + [$_M$ Past], 2

The formulation in (68) assumes a [$_V$ V – V] analysis of compound verbs as in Dasgupta (1977). The need for V′ has been eliminated by formalizing Inflection the way we have. As in certain conceptions of the application of iterable phonological rules, (68) will operate in such a way as to yield only left-branching structures, not right-branching structures or structures involving a mixture of left and right branching.

So far, for ease of exposition, I have assumed that all Inflection Rules apply to the empty initial tree after phrase structure rules have finished forming it and before lexical insertion has put words in it. However, in view of the discussion of Phrase Formation in chapter 3, one may maintain that Inflection Rules that apply inside compound verbs belong in the lexical rather than the syntactic component of the grammar. Specifically, I will assume that the following sequence of procedures operates on compound verbs.

(69) Step one: Apply the Phrase Formation Rule V ➔ V V (V being one of the lexicon's many initial symbols):
[$_V$ [$_V$][$_V$]]

Step two: Apply the lexical Inflection Rule (68) to the output of step one
[$_V$ [$_V$ [$_V$][$_M$ Past]] [$_V$]]

Step three: Insert a rudimentary compound verb lexical entry into the V slots:
[$_V$ [$_V$ [$_V$ /bOS/][$_M$ Past]] [$_V$ /pOR/]]

The output of step three of (69) is a 'lexetic representation', a full-fledged lexical entry which may now seek to enter a V slot in a phrase-marker which has, after coming out of the PS component, been further processed by Syntactic Inflection Rules like (63)-(65) so that the

prospective compound verb occupant of the main verb slot of a particular clause may check to see if the T, PERS, HON put in by the Inflection Rules are compatible with the compound verb lexical entry's particular requirements. In yet another respect, then, we have now outgrown the V′ analysis of chapter 3.

Phrase Structure rules and Inflection rules are both parcelled out now between the lexical (irregularity-tolerant) and syntactic (regularity-insistent) parts of the grammar. Rules applicable to or within the subtree dominated by a lexical node such as V are now assumed to belong to the lexicon. Rules pertaining to bigger domains belong to the syntax. To bring this out clearly, let us simplify the terminology. Scrapping the fancy term Phrase Formation Rule, we will speak of an orthogonal relation between two sorts of rules – phrase-structural and inflectional – and two allegiances – to the syntax or to the lexicon.

(70)	belongs to syntax	belongs to lexicon
PS Rules	VP → NP V, etc.	V → V V, etc.
Inflection Rules	Affix T to V]$_S$, etc.	Affix [$_M$ Past] to V in [$_V$ __V], etc.

The particular operations that we had been calling just "Phrase Structure Rules" can now be specified as "Phrase Structure Rules belonging to the lexicon".

In this system, the problem of infinitive inflection assignment and conjunctive inflection assignment applying to the same stem and yielding monstrosities like (59) simply does not arise.

We turn now to negation. Given the division of labour between the lexicon and the syntax, syntactic inflection rules will never look into a lexical node like V. Thus, when analyzing a structure like [$_S$ ...[$_V$ V[$_V$ V V]]], a syntactic inflection rule adjoining NEG to V will only see the maximal V. This is what we want. We want to prevent more than one /na/ making its way into a compound verb. So, we posit a syntactic inflection rule (71) and order it after (65).

(71) S. D. : [$_S$ X – V – Y]
S. C. : 1 2 3 ➔ 1, 2+NEG, 3

Like other inflection rules, (71) should not be allowed to apply to its own output. There will have to be a general convention on inflection rules to this end.

Udaya Narayana Singh's (1976) exhaustive account of the distribution of pre- and post-verbal negatives in Bangla suggests that it might be better to postulate two inflection rules, (72) and (73), than to have a single suffixing rule (71) plus a movement rule interchanging V and NEG in certain environments.

(72) S. D. : [$_{[-S]}$ X – V]
S. C. : 1 2 ➔ 1, NEG+2

(73) S. D. : [$_S$ X – V]
S. C. : 1 2 ➔ 1, 2+NEG

These formulations reflect a pattern that emerges from Singh's data: the main verb of a clause

takes a post-verbal negative, while non-S construction put the negative before the verb. Notice that the symbol [–S] in (72), a bar-notional symbol for the natural class of all nodes that are not S or S′, makes considerably more sense in Norbert Hornstein's (1975) system of bar notation than in David Halitsky's (1975). (72) needs to follow (67) (and (68), but that is taken care of, because (68) is now a lexical inflection rule and will precede (72) as a matter of course). (73) must follow (65).

We now have an account of the contrast between the modal construction's tolerance for double negation and the compound verb's intolerance. The compound verb as a whole will be allowed one helping of a NEG-attaching infection rule. If the compound verb is the main verb of a clause, the relevant rule will be (73); otherwise, (72). So, a compound verb will never get two negatives. But, in a modal construction, the modal, being the main verb of a clause or of a VP, is entitled to its own NEG (by rule (73) or (72), respectively), and so is the infinitival complement's main verb, by rule (72). One can therefore easily end up with two negations, one for each of the verbs.

As promised, the present section has solved (59) and answered the deeper question (why modal constructions, though they resemble compound verbs in surface structure, differ from them otherwise) by revising the analysis reasonably thoroughly without reversing the headway made in section 4.8. The mechanisms introduced in the present section work at an early point in the derivation, between the point at which PS rules leave off and the point at which transformations begin; the results of section 4.8 has to do with goings-on in the transformational subcomponent of the syntactic component, which remain unaffected by section 4.9.

4.10 More about infinitives and gerunds

Now that we know where infinitives stand, we return to our inquiry into how and why gerunds block them in various positions. Recall:

(74) Subject-laden Gerund blocks subject-laden infinitive in subject position
Subject-laden Gerund blocks subject-laden infinitive in complement position
Subjectless Gerund blocks subjectless infinitive in subject position

If our analysis of infinitive phrases is correct, then the reason why subject-laden or subjectless gerunds block subject-laden or subjectless infinitives in subject position is that, in the position of Subject of an S, a gerund phrase, being an NP, may occur, thanks to the rule S → NP ... V, whereas an infinitive phrase, being a VP, will not occur, since there is no rule S → VP ... V. This leaves unexplained the second line of (74) – why we get no subject-laden infinitives in complement position, while we do get subject-laden gerunds in complement position. Again, the gerunds can be there because the rule expanding S makes room for an NP object of the main V of the S. This NP, like any other NP, is free to be a gerund phrase that contains a subject. The reason for the non-occurrence of subject-laden infinitives in complement position, or anywhere else, is just that infinitive phrases, being VPs, never have subjects.

All this may sound awfully anti-climactic after the great to-do about blocking. The real point one might make about this case of blocking would be a functional and historical one.

Bangla developed its system of gerunds in such a way that it never 'needed' to develop its infinitival structures away from mere VP-hood towards a fuller, more gerund-phrase-like distribution the way English has. To some extent the shift from the infinitival VP system to the infinitival S system is still under way in English, leaving amply confusing data for parties to the Chomsky-Bresnan debate on this matter to use in grinding their respective axes. But in Bangla the Bresnan side gets a walkover. It is deafeningly clear that infinitival structures in Bangla are always VP and never S, that they never have subjects. The notion of untensed S which Chomsky has postulated for English, with some success, has no content in Bangla, as will be shown in chapter 5, among other things.

The main task of chapter 5 will be to finish describing the verb system and to relate it to clause structure.

Chapter 5

VERBS AND CLAUSE STRUCTURE

5.0 Strategy and tactics

5.0.1 Strategy

As the absence of the words 'overall Bangla grammar' in the chapter title indicates, I propose to move back towards the main issues of this volume which were adumbrated in chapter 1, now that the reader is familiar with some basic facts about Bangla syntax and morphology.

This dissertation concerns itself mainly with certain types of clauses. It seems appropriate to approach the study of clause types negatively first, by showing that, contrary to universal analyses proposed by Chomsky & Lasnik (1977), Chomsky (1978), and other authors, infinitive structures in at least one language (Bangla) are not embedded clauses, a fact which enables us to explain why infinitive structures in Bangla do not function as relative clauses or constituent questions. A discussion of the non-clausality of infinitive structures must, for the sake of completeness, keep as much as possible of the syntax of various verb forms in view – including some facts about Bangla verbs which did not get mentioned in chapters 3 and 4.

5.0.2 Tactics

This chapter makes its point first and fills in the background later. In order to make the strongest possible case for the non-clausal status of infinitive structures in Bangla, I concede, for argument's sake, Chomsky's (1978) thesis that every infinitive structure in English is a clause (an S′ in his analysis). I then show that, even if one accepts the arguments which, for some linguists, establish the clausal status of the infinitive structure in English, it still remains necessary to analyze Bangla infinitive structures as other than clausal. If one makes other assumptions about English, then of course one does not expect Bangla infinitive constructions to be clausal in the first place. That would make the following demonstration as redundant as, say, a demonstration that Bangla nasalized vowels were non-implosive. So, I find it rhetorically more useful to concentrate on readers who, like Chomsky and unlike Bresnan, believe that all English infinitive constructions are clausal and who therefore need to be persuaded to make an exception for Bangla.

After showing that Bangla infinitive structures are non-clausal, I put this matter in perspective by discussing aspects of the syntax of Bangla verbs that did not find their way into earlier chapters. Then I return to the non-clausality of infinitive structures and on its basis make and test the prediction that Bangla infinitive structures do not do the thing that clauses do which interests us – that they do not function as relative clauses or constituent questions.

5.1 The non-clausality of Bangla infinitive structures

Chapter 4 hypothesized that Bangla infinitive structures were of the form [$_{VP}$ (NP) (VP) V] and that unlike gerunds they could never contain a subject for their head V. This analysis must now be compared with a natural alternative which, mimicking Chomsky and Lasnik's (1977) and Chomsky's (1978) proposals for English, analyzes Bangla infinitive structures as in (1). For ease of comparison, my hypothesis is set forth in (2).

(1) **CIH**: the Clausal Infinitive Hypothesis.

a. Bangla has no VP node at all.

b. Infinitive structures are of the form [$_{S}$ NP (NP) (S) V].

c. Somehow, the group of Inflection Rules affixing T PERS (HON) to verbs is *en bloc* optional, so that any S-final V may be tensed or untensed.

d. A filter rejects (i) root sentences without a tensed main V, (ii) subject clauses without a tensed V.

e. Lexical insertion transformations are optional, and the nodes they leave empty may remain unfilled even at surface structure.

f. Depending on the lexical dictates of the matrix V in ..[$_{S}$NP..] V.., the NP subject of the infinitival complement, (i) if [$_{NP}$ *e*], undergoes coindexing with an upstairs NP specified by the matrix V; (ii) otherwise, must be assigned Objective Case.

(2) **NCIH**: the Non-Clausal Infinitive Hypothesis.

a. Bangla has a VP node whose head V is assigned Infinitive by an Inflection Rule.

b. Infinitive structures are of the form [$_{VP}$ (NP) (VP) V].

c. Because of the formulation of Inflection Rules, all and only S-final Vs are tensed.

d. Root sentences or subject clauses without a tensed main V do not arise. So, no filter.

e. Lexical insertion transformations are optional, but a filter rejects empty non-maximal phrase nodes. (See chapter 3.)

f. There is no 'subject of the Infinitive' that would have been eligible to undergo matrix-determined Case assignment or coindexing. Thus, these operations do not take place. But there is the ordinary Case assignment of NP$_{S}$ by their right sister V, partially determined by lexical properties of the V — by its subcategorization features in the sense of Chomsky (1965).

My first argument supporting NCIH against CIH rests on the Inflection Rules introducing NEG. These rules emerge from Singh's (1976) work on the distribution of NEG.

(IV-72) S.D.: $[_{[-S]}$ X – V]
S.C.: 1 2 ➔ 1, NEG+2

(IV-73) S.D.: $[_{S}$ X – V]
S.C.: 1 2 ➔ 1, 2+NEG

Chapter 3 showed that gerund phrases and participle phrases are not clauses. Thus, the fact that gerunds and participles take negatives according to (IV-72) while bona fide clauses take negatives according to (IV-73) provides primary motivation for these rules which remains unaffected even if, for argument's sake, one grants the possibility that Bangla infinitive structures may be clausal. Independent motivation for the ±S feature is given by Hornstein (1975). He shows that the S-S′ system must be separated from the V-bar system. With respect to the NEG-adjoining Inflection rules, the difference between CIH and NCIH is that NCIH plus these NEG rules correctly predicts that NEG precedes the infinitive whereas CIH in conjunction with rule (IV-73) makes the wrong prediction – that the infinitive precedes NEG. Thus, NCIH is superior to CIH.

My second argument again pertains to the difference between S and the V-bar system, and again relies on Hornstein (1975) for independent support. Chapter 4 showed that, when the head V of an infinitive structure moves out by Infinitive Reanalysis to left-adjoin to the matrix verb, the VP which the head V was holding up collapses, is pruned. This result, independently obtained by Schwartz (1972), is consistent with NCIH. But adoption of CIH forces one to postulate that an S structure collapses when its head V is removed. Within the bar-notational framework this would make S part of the V-bar system. But S is not part of this system: see Hornstein and (IV-72, 73) above. Therefore NCIH is superior to CIH.

My third argument attacks (1-f), specifically, the part about obligatory assignment of Objective Case to the non-empty NP subject of an infinitive. The facts that (1-f) is designed to handle are as follows. For the perception verbs /dekh/ 'see' and /Sun/ 'hear', a proponent of CIH will need to claim that their lexical entries dictate assignment of Objective Case to the non-empty NP subject of their infinitival complement in underlying structures like (3)-(4), yielding (5)-(6).

(3) ami $[_{S}[_{NP}$ ram] jete] dekhlam "I Ram to-go saw"

(4) ami $[_{S}[_{NP}$ SEm] gayte] Sunlam "I Shyam to-sing heard"

(5) ami ramke jete dekhlam 'I saw Ram go'

(6) ami SEmke gayte Sunlam 'I heard Shyam sing'

In contrast (a proponent of CIH will add) the verb /cah/ 'want' has a lexical entry which induces coindexing of the empty NP subject of its infinitival complement with the subject of /cah/ in surface structures like (7), yielding (8). Since /cah/ does not assign Objective Case in its infinitival complement, *ram* in the structure (9) will fail to receive Case and will be rejected,

correctly, by a filter prohibiting uncased lexical nouns.

(7) ami [$_S$ [$_{NP}$ *e*] jitte cay “I *e* win want” ‘I want to win’.

(8) [$_{NPi}$ ami][$_S$ [$_{NPi}$ *e*] jitte] cay “I_i e_i win want” ‘I want to win’

(9) *ami [$_S$ [$_{NP}$ ram][jitte] cay “I Ram win want” ‘I want Ram to win’

My argument against such a proposal is based on the following:

(10) ami ramke dekhlam “I Ram-OBJ saw” ‘I saw Ram’

(11) *ami ram jete dekhlam “I Ram go saw”

(12) *ami ram dekhlam “I Ram saw”

(13) ami pata jhorte dekhlam “I leaf fall saw” ‘I saw leaves fall’

(14) *ami pata dekhlam “I leaf saw” ‘I saw leaves’

(15) *ami patake jhorte dekhlam “I leaf-OBJ fall saw”

(16) *ami patake dekhlam “I leaf-OBJ saw”

(17) ami manuS dekhlam “I person saw” ‘I saw people’

(18) ami manuS morte dekhlam “I person die saw” ‘I saw people die’

(19) ami manuSke morte dekhlam “I person-OBJ die saw” ‘I saw human beings die’

(20) ?ami manuSke dekhlam “I person-OBJ saw” ‘I saw Man’

(21) ami manuSke biSSaS kori “I person-OBJ belief do” ‘I believe people’

(22) *ami manuS biSSaS kori “I person belief do”

(23) *ami manuSke Sunechi “I person-OBJ have-heard”

(24) ami manuSke katrate Sunechi “I person-OBJ groan have-heard” ‘I have heard people groan’

(25) *ami manuS Sunechi “I person have-heard”

(26) *ami manuS katrate Sunechi “I person groan have-heard”.

Initially, on the basis of (5) and (10)-(16), one may be inclined to argue that, regardless of how

Objective Case was assigned to a particular NP, the surface manifestation of Objective once assigned is a property of the head N. If the head N is inanimate it will show no overt marker; if human and/or proper it will take a Case ending, /ke/ in the singular.

However, this initial conclusion is contradicted at once by (17)-(26). (17)-(20) show, first of all, that some nouns vary as to whether they show an overt Objective marker, and, secondly, that the variation is significant and appears to be linked with the syntactic structure of the sentence. Thus, with the noun /manuS/ it is harder to get the form /manuS-ke/ in the environment /ami ___ dekhlam/ than in the environment /ami ___ morte dekhlam/. (20) has only a literary, metaphoric interpretation; one may wish to exclude (20) for the purposes of straight grammatical description. If /dekh/ 'see' apparently resists taking *manuSke* as its left neighbour, (21)-(22) indicate that the verb /biSSaS kOr/ 'believe' on the contrary requires *manuSke* rather than *manuS* as its left neighbour – and, since /biSSaS kOr/ takes no infinitival complements, the question of the environment /ami ___ V biSSaS kori/ does not arise. There must be markings in the lexical entries for /dekh/ and /biSSaS kOr/, and perhaps also for /manuS/, registering these idiosyncrasies. (23)-(26) show that the verb /Sun/ 'hear' never takes *manuS(ke)* as an immediate left sister at all, and permits only *manuSke* (not *manuS*) in the environment /ami ___V Sunechi/; again, these facts about /Sun/ must appear in /Sun/'s lexical entry.

The data show that different verbs have different lexical requirements of presence and absence of Objective Case marking in the environment ___V and the environment ___V V. Chomsky (1965), supported in this respect by later research, correlates the existence of such lexical selection with the existence of a bidirectional Grammatical Relation – a Selectional Relation. In this case the data must thus be regarded as indicating that in all the well-formed examples in (10)-(26) the NP in the Objective is in a Selectional Relation of "Object-Main-Verb" with the V which so complexly determines its Case assignment. (1) is just too simple to cope with the data. (2), which unlike (1) makes the use of selectional features possible in both the pre-infinitival NP examples and the immediate sister NP examples, must be adopted as the right analysis of infinitive structures.

My fourth argument supporting NCIH invokes Occam's Razor.

Despite the dearth of studies on the subject, presumably careful semantic analysis of English would show that, at least in terms of truth value patterns, propositional structure, and argument positioning, the interpretation of (27) does not differ from that of (28).

(27) He can row

(28) He is able to row

Analyses of English which correspond to CIH for Bangla assign respectively the following construed surface structures to the above sentences.

(29) $[_S\,[_{NP}$ he] can row

(30) $[_S\,[_{NPi}$ he] is able $[_S\,[_{NPi}\,e]$ to row]]

Structure (29) has only a single subject, *he*, for the verb *row* as well as the modal *can*. The modal *can*, like other modals, must of course be regarded as a verb for the purposes of

Affix Hopping and, as the semantic parity between (27) and (28) indicates, of predicate interpretation. While there is debate about the correctness of (30) – see Brame (1976) for a critique of (30) – no one doubts that (29) is, in its essentials, the right surface structure for (27). Linguists agree that, whatever the mechanism of semantic interpretation may be, they must be able to process (29) without the special sort of aid that (30) provides – without, that is, two coindexed subject nodes, one for each predicate expression.

Since the apparatus for interpreting a single NP, like *he* in (29), in relation to two predicate expressions, like *can* and *row* in (29), is available, and must be able to function even though an NP in one phrase-marker will play 'Subject-of' just one V, it follows that the input to rules determining the argument structure of (31) may have the form (33), as proposed by NCIH, and need not have the form (32) required by CIH.

(31) ram jete pabe "Ram go will-get" 'Ram will get to go'

(32) $[_{NPi}$ ram$][_S [_{NPi}\ e]$ jete] pabe

(33) $[_{NP}$ ram$][_{VP}$ jete$][_V$ pabe], deep structure; or, $[_{NP}$ ram$][_V [_V$ jete$][_V$ pabe], surface structure; whichever is input to semantics.

Since the foregoing arguments have shown that there is no syntactic motivation for the analysis which yields such representations as (32), the only reason one might still have for such representations is the belief that the semantics of such sentences works better from (32) than from (33). But our consideration of (29) and (30) shows this belief to be false. The NP *ram* in (33) is the syntactic subject of only one verb, *pabe* (which agrees with *ram* for Person and Honorificity); nevertheless, semantic interpretation will be able to show the relations that *ram* bears to both *jete* and *pabe* without difficulty.

The four arguments just given suffice to refute (1), the Clausal Infinitive Hypothesis. It can be shown that the arguments also apply against other versions of CIH which one might build by mimicking analyses of English infinitive structures which were current in the sixties – analyses in term of deletion rather than interpretation. I will not try to carry out this task, whose steps seem obvious enough.

The fourth argument is the only one with any significant analogue in English. But notice that this argument has no force on its own but relies on a prior demolition of syntactic motivation for the clausal analysis. Thus, the result obtained for Bangla leaves the debate about the analysis of English infinitives more or less where it was. In chapter 10 I offer some fresh contributions to that debate.

5.2 The 'Passive'

Since I brought up grammatical relations in my third and fourth arguments, some readers are likely to want to know about 'the Bangla passive', hoping to find interactions which will help test the available options. This is therefore the right place to elaborate on my comments in section 3.2.3 about how Bangla really lacks a passive but can variously mimic foreign passives. I will concentrate here on the counterpart to the English passive. The /ja/-construction which mimics the Hindi passive has, for our purposes, the same properties as the

/hO/-construction which mimics the English passive.

(34) $[_S [_{NP}$ bhat$][_V [_N [_V$ kha] Wa$][_V$ hObe]]] "rice eating will-happen"

Sentences like (34) invite translation as 'Rice will be eaten'. In fact, thanks to the coincidence of the gerund $[_N$ khaWa] with the participle $[_A$ khaWa], the temptation is to use a word-for-gloss like "rice eaten will-be", *hO* being a stem which may mean 'be' or 'become' or 'happen', In order to ascertain whether we should take that option or reject it, we turn to examples of gerunds which do not function as participles, such as *dowRono* 'running'. It turns out that the initial constituent of the putative passive form is a gerund rather than a participle, since *dowRono* does occur here.

(35) $[_S [_V [_N [_V$ dowRo] no$][_V$ hObe]]]"running will-happen"

(35) cannot be translated with a passive sentence in English, though it can in German: *Es wird gerannt werden*. The question of whether there may be some virtue to calling (35) a passive sentence, while perhaps not trivial or pointless, has no bearing on our concerns here. We want to know whether these putative passives change grammatical relations. The answer to that question becomes obvious from the evidence in (36)-(39).

(36) dhopake kapoR daW "washerman-OBJ clothes give" 'Give the washerman clothes'

(37) dhopake kapoR deWa hok "washerman-OBJ clothes giving happen-IMP" 'Let clothes be given to the washerman'

(38) montrider dillite paThabo "ministers-OBJ to-Delhi we-will-send" 'We will send the ministers to Delhi'.

(39) montrider dillite paThano hObe "ministers-OBJ to-Delhi sending will-happen" 'The ministers will be sent to Delhi'

The answer is that the same markings occur in the 'passive' as in the 'active', at least for the indirect and direct objects. The subject in the 'active' usually cannot be present in the 'passive', as in (40) where the sentence is ill-formed with and well-formed without the subject, but when the NP is retained it takes a genitive ending in the 'passive' sentence, as in (41). This genitive is explicable in terms of the structure of gerund phrases; see (42). The 'subject of the active sentence' counts, in (41), as semantically speaking the subject of a gerund phrase where the gerund is the head noun; recall that the subject of a gerund phrase is normally in the genitive.

(40) (*bimOler) Sumitke ThEngano hoto "(Bimal's) Shumit-OBJ thrashing used-to-happen" 'Shumit used to be thrashed (by Bimal)'

(41) bimOler bhat khaWa hObe "Bimal's rice eating will-happen" 'Rice will be eaten by Bimal'

(42) As in R.C. Dougherty's unpublished gestalt comparison analysis of English *if-then*

structures, one may propose that subject-retaining Bangla 'passive' sentences like (41) have their syntax and semantics interrelated as follows. On the one hand, (41) is generated with the syntactic structure NP NP [$_V$N V], the genitive form for the first NP being selected at random from the language's Case repertoire. On the other hand, independently, a second structure [$_{NP}$ NP NP N] V is generated with the NP having the internal structure of a gerund phrase as described in chapter 3. The first structure is matched with the second to provide a joint description of the properties of the string. The NP NP [$_V$ N V] structure gives the constituent breaks etc. The [$_{NP}$ NP NP N] V structure determines the grammatical relations and propositional and argument structure. Some members of the first set of structures, e.g. *bimOl bhat khaWa hObe* "Bimal rice eating will-Copula", will fail to match up with any member of the second set and thus be marked as ill-formed. Some members of the second set, e.g. *bimOler Sumitke ThEngano hoto* "Bimal's Sumit-OBJ beating used-to-Copula", will fail to match up with any member of the first set and thus be marked as ill-formed. Only strings which can be parsed both ways are well-formed.

Turning to the interaction of all this with the infinitive, we find that the 'passive' construction has nothing to tell us about the grammatical relations over and above what we ascertained in section 5.1:

(43) ramke jete dEkha gElo "Ram-OBJ go seeing went" 'Ram was seen going'

If we set aside the fact that, for selectional reasons, (43) requires the /ja/ "go" form rather than the /hO/ "happen" form, (43) is an ordinary 'passive' comparable to the examples we have been looking at, and comparable to (44), which has no infinitival structure in it.

(44) ramke dEkha gElo "Ram-OBJ seeing went" 'Ram was seen'

The complexities with *manuS* which were crucial in my third argument against CIH turn up again, unchanged, in the 'passive'.

(45) manuS dEkha gElo "person seeing went" 'people came into view'

(46) manuS morte dEkha gElo "person die seeing-went" 'People were seen dying'

(47) manuSke morte dEkha gElo "person-OBJ die seeing went" 'Human beings were seen dying'

(48) ?manuSke dEkha gElo "person-OBJ seeing went" 'Behold, Man came into view'

The function of this section was the negative one of convincing some readers that in Bangla they have nothing to learn about grammatical relations by looking at passive structures, for there really are no passive structures in the sense in which the term is used for other languages. I turn now to the causative with a similar end in view. These sections also serve to introduce aspects of the verb system which would otherwise remain undescribed, leaving gaps in the outline grammar.

5.3 The Causative

As the lack of quotes in the title of the section shows, I believe that Bangla does have causative verbs. (49) shows some non-causative stems whose causative counterparts are given in (50).

(49) jhOr 'shed', pOR 'read, study', Son 'hear, listen', dEkh 'see'

(50) jhOra '(cause to) shed', pORa 'teach, cause to read/study', Sona 'play, sing', dEkha 'show'

I will now present an analysis of causative morphology in Bangla based on my paper submitted for publication in *A Festschrift for Robert Fowkes.*[7]

Chapter 2 stated the following rule without discussion.

(II-13-2b) Lowering

{i, u} → {e, o}/___C_0<Wa>]$_{verb}$: <complexly optional>

Lowering, as a normal phonological rule, is obligatory except where otherwise stated. The angled brackets notation informs us that its ___$C_0]_V$ case is obligatory, while its ___C_0Wa]$_V$ case is 'complexly optional'. By that cryptic phrase I meant the following. The second case of the rule applies optionally to /u/ but never applies to /i/. Thus, /ulT-Wa/ shows a surface first syllable alternation of *u* and *o*, but /pichl-Wa/ shows no surface alternation of *i* and *e*.

Now consider *Son* and *Sona* in (49)-(50). The obligatory case of lowering derives *Son* from /Sun/. One would expect the optional subcase to yield a surface alternation for /Sun-Wa/. *Sona* should have an alternant beginning with *Su-*. But it does not. Causative stems always undergo Lowering.

One concludes that the underlying form of causatives must be not [$_V$..Wa] but [$_V$[$_V$..]Wa], so that they attract the obligatory case of Lowering. /Sun-Wa/ is [$_V$[$_V$Sun]Wa], then, in contrast to /ulT-Wa/ which is [$_V$ulT-Wa]. However, stems such as /ulT-Wa/ which look like but are not causatives must still be considered bimorphemic. Their /Wa/ must be equated with the /Wa/ which forms causatives, at least for the purposes of the statement that, whenever /Wa/ immediately precedes the participle-gerund ending, this ending selects the allomorph /no/ instead of /Wa/. This statement is true regardless of the function, causativizing or otherwise, of the /Wa/ which precedes the participle-gerund ending.

In order to ensure proper application of the two cases of Lowering one apparently must violate A-over-A. Thus (51) must undergo the /Wa/ case of Lowering. (52) attracts instead the ___$C_0]_V$ case. The only principle which would guarantee such application to (51)-(52) would be some principle which would always cause the lower V node (if two V nodes are present) to undergo the rule – a sort of A-under-A principle if V is taken to be the target of the rule.

---

[7] Editorial note, 2020: That festschrift plan came to nought. The text later appeared in a different festschrift (Dasgupta 2000); a reanalysis of the data in terms of a new framework appeared as Dasgupta (2001).

(51) [$_V$bul-Wa]: bulo, bola ‘stroke’

(52) [$_V$[$_V$bhul]Wa]: bhola ‘comfort, cause to forget’; *bhulo

One solution to this problem would be to regard V not as the target but as the domain of the rule. Then the A-over-A principle itself, or rather a proviso on it given in *Language and Mind* (Chomsky 1972), picks out the minimal domain, which is the result we want.

Alternatively, if the outer V of forms like (52) gets processed only on a relatively late cycle, one might propose that Lowering is not available on that cycle. Distinguishing between roots like (51) and stems like (52), one might then treat Lowering and rules crucially preceding it as root-level rules of phonology. Since the only other rule in (II-13) making explicit reference to]$_{Verb}$, a rule which (II-13) places later in the ordering than is strictly necessary, is (5b) –

(II-13-5b) Progressive Raising

$$a \rightarrow o \; / \; \{ i, y, u, w \} \; C_0 \_\_\_]_{Verb}$$

– a rule which also always affects roots and never stems, it becomes possible, then, to propose that rules of Bangla phonology that need to mention the V-bracket belong to the root-level set of rules that precedes the stem cycle.

Returning to more word-external grammatical concerns, we ask for evidence which will select, for Bangla, one of the alternatives analogous to the following options which Chomsky considered for *John grows tomatoes*.

(53) John [+cause][$_S$ tomatoes grow]

(54) John [+cause, grow] tomatoes

In his exegesis of (54) Chomsky (1970: 215) postulated ‘that there is a feature [+cause] which can be assigned to certain verbs as a lexical property. Associated with this feature are certain redundancy rules which are, in this case, universal, hence not part of the grammar of English but rather among the principles by which any grammar is interpreted. These principles specify that an intransitive with the feature [+cause] becomes transitive and that its selectional features are systematically revised so that the former subject becomes the object. Similar principles of redundancy apply to the associated rules of semantic interpretation.’

Crucial evidence for preferring, for Bangla, an analysis of causatives more in line with (54) than with (53) comes from causative stems like /jal-Wa/ derived from /jOl/. Consider first (55), which shows the normal use of these stems. (For some speakers, for whom (55-ii) must have *jallo* and not *jalalo*, one faces the extra problem of accounting for the absence of the causative morph *-a-* in this word.)

(55) i. agun jollo “fire lit” ‘A fire was lit/kindled’
ii. o agun jalalo “(s)he fire lit” ‘(S)he lit a fire’

Now, the causative stem /jal-Wa/ also has an entirely acceptable and natural use, which one

must consider figurative in some sense.

(56) tumi oke jalaccho “you him/her are bothering”
‘You are bothering him/her’

In contrast, the use of the intransitive /jOl/ in this figurative sense sounds quite unnatural, at best like a clumsy attempt to be colourful or poetic or funny. It sounds, in fact, like a back-formation from (56).

(57) *o jolche “(s)he is-‘burning’”

This semantic contrast between (56) and (57) indicates that /jal-Wa / is an independent lexical item with properties of its own that are only in part explainable in terms of /jOl/ and the predicate-raising transformation which analysis (53) would use in deriving causatives from non-causatives. Notice also the vowel difference between the roots – synchronically an opaque difference. This difference calls for an explanation (unavailable on the synchronic plane) if one derives the causative transformationally from the non-causative. Thus, an analysis of the (54) type, a lexical analysis, is better than an analysis of the (53) type, a transformational analysis.

To clinch the issue, here is some more evidence – less neat to my mind though perhaps more persuasive for others.

(58) cheler poSak chaRaW “son’s clothing cause-to-doff-IMP” ‘Undress (your) son’

(59) chele poSak chaRbe “son clothing will-doff” ‘The son will undress’

(60) peMajer khoSa chaRaW “onion-GEN skin cause-to-doff-IMP” ‘Skin the onion’

(61) *peMaj khoSa chaRbe “onion skin will-doff” ‘The onion will skin’

(62) rOmeSbabur kache chatrora ingreji pORe “Romesh-Babu-GEN at students English study” ‘The students learn English from Romesh-Babu’

(63) rOmeSbabu (chatroder) (ingreji) pORan “Romesh-Babu (students-OBJ) (English) teaches” ‘Romesh-Babu teaches (the students) English’

(64) *rOmeSbabur kache chatro (ingreji) pORe “Romesh-Babu-GEN at student (English) studies”

(65) rOmeSbabu chatro (*ingreji) pORan (OK without *ingreji*, * with *ingreji*) “Romesh-Babu student (*English) teaches” ‘Romesh-Babu teaches students (*English)’

(66) jhunu birokto hOY “Jhunu irritated gets” ‘Jhunu gets irritated’

(67) tapoS jhunuke birokto kOre “Taposh Jhunu-OBJ irritated makes” ‘Taposh irritates Jhunu’

(68) tapoS Taka paY “Taposh money gets” ‘Taposh gets money’

(69) tOpeS tapoS ke Taka dEY “Topesh Taposh-OBJ money gives” ‘Topesh gives Taposh money’

(70) tapoS Sahajjo paY “Taposh help gets” ‘Taposh gets help’

(71) ?jhunu tapoSke Sahajjo dEY “Jhunu Taposh-OBJ help gives” ‘Jhunu gives Taposh assistance’ (sounds like a bad translation from English)

(72) jhunu tapoSke Sahajjo kOre “Jhunu Taposh-OBJ help does” ‘Jhunu helps Taposh’

The /chaR/-/chaR-Wa/ evidence, (58)-(61), resembles the evidence given for /jOl/ and /jal-Wa/, and the argument is the same, except that in this case there is no phonological discrepancy between the nuclei to explain.

The next data set is (62)-(65), which harks back to the discussion of the distribution of the Case marker *-ke* in section 5.1. In terms of the lexical hypothesis (54), (63) is taken to show that /pOR-Wa/ normally takes two objects and assigns *-ke* to its indirect object, but, as (65) shows, /pOR-Wa/ does have a use which enables it to take an unmarked ‘indirect’ object (this structure does not tolerate a direct object, hence the “(**ingreji*)”, indicating that the structure is unacceptable if the word *ingreji* is present there). (62) and (64) show that a use of /pOR/ exists corresponding to the normal, two-object use of /pOR-Wa/, but no use of /pOR/ exists corresponding to the special, unsuffixed-direct-object-only use of /pOR-Wa/. Again, the transformational view (53) counterfactually predicts the non-existence of such discrepancies between the causative stem and its base.

(66)-(67) and (68)-(69) are pairs of sentences which, because of the semantic and selectional relation between them, the transformational hypothesis would have to analyze in terms of causative derivation in order not to miss the generalization. But (70)-(72) would refute any analysis of (66)-(69) in terms of a regular derivation of /kOr/ from /hO/ and of /di/ from /pa/. So a transformationalist would be compelled to leave /kOr/-/hO/ and /di/-/pa/ alone and therefore also to devise interpretive and lexical apparatus for expressing the partly systematic interrelation between the members of these pairs. This apparatus would perforce be available for the likewise not fully systematic relation between V and V-*Wa*, leaving the burden of proof on the proponent of the transformational hypothesis.

As in section 5.2, grammarians who hoped that causative structures would illuminate grammatical relations and perhaps lead to ways of disproving the analysis of verb phrases and sentences developed here have been disappointed. Bangla has no ‘causative structures’ of the sort that many modern grammarians might seek, but only causative verbs bearing a not entirely consistent relation to their non-causative bases.

An obvious question for an unregenerate anti-lexicalist to ask at this point is: don’t causative verbs have an ability (denied to ordinary verbs, which are shackled by strict subcategorization features limiting them to their own clause) to assign a second overtly marked Objective Case? How can Bangla express the interpretation (73), for instance, if not by means of the expression (74)?

(73) He made Esha give Uncle the news

(74) [$_S$[$_{NP}$ o][$_S$[$_{NP}$ eSake][$_{NP}$ meSoke] [$_{NP}$khOborTa][$_V$de-]][$_V$-Wa-lo]] “(s)he Esha-OBJ

Uncle-OBJ news-item give-CAUSE-PAST"

The answer is that (74), with its successive NPs marked with the Objective marker /ke/, is a borderline sentence of the sort native speakers try and plan their way around, like German (75) and Bangla (76). Since at least Bangla (76), which is analogous to (74), has nothing to do with the causative system (there being no non-causative verb corresponding to /bec/ 'sell'), it seems that we are simply looking at the effects of an anti-repetition constraint similar to the one against the double Infinitive form /kOr-yte thak-yte/ – see the discussion following the chart in section 2.3.2. In any case, there is no reason to prefer (74) to (77) as the appropriate structure for the borderline sentence.

(75) ?Ich will ihn ihn schlagen sehen "I want him him hit see"
'I want to see him hit him'

(76) ?ami tomake biruke becbo "I you-OBJ Biru-OBJ will-sell"
'I will sell you Biru' (where Biru is, say, a puppy)

(77) [$_{S}$[$_{NP}$ o][$_{NP}$ eSake][$_{NP}$ meSoke] [$_{NP}$khOborTa][$_{V}$deWalo]]

One last point about causative verbs. This is meant less as a parting salvo against the transformational hypothesis than as a way to relate the subject to the discussion of passives that went before. Two causative stems, /dekh-Wa/ and /Sun-Wa/, derived from /dekh/ 'see' and /Sun/ 'hear', respectively, have not only the causative meanings 'show' and 'sing, play,' but also the anti-causative or 'passive' meanings 'look' and 'sound':

(78) biruke khaSa dEkhacche "Biru-OBJ great is-looking" 'Biru looks great'

(79) ganTa Opurbo Sonalo "song-item fantastic sounded" 'The song sounded fantastic'

Like other 'passive' constructions, this one also, as the undisturbed /ke/ on /biruke/ indicates, leaves its Objects alone. A historical account of this exceptional use of /dekh-Wa/ and /Sun-Wa/ can be given. Synchronically, however, there is as little reason to set up a separate pair of lexical items /dekh-Wa/ and /Sun-Wa/ for the 'passive' use as there would be for distinguishing *ears* of corn as a separate lexical item from human or animal *ears* because of the once evident lexical difference.

5.4 Aspect and Negation

The perfect aspect 'indicative present' and 'past' forms of the verb /kOr/ given in chapter 2 are repeated below with the slight modification that the word beginning with *ch* is in every case suffixed to the preceding word, in deference to traditional orthographic habits.

(80)	1p.	2p. int.	2p. neutr.	3p. neutr.	non-1p. hon.
PRES	korechi	korechiS	korecho	koreche	korechen

PAST korechilam korechili korechile korechilo korechilen

One would expect the corresponding negative forms to be *korechi na, korechili na,* etc. What one gets instead is (81).

(81) kori ni koriS ni kOro ni kOre ni kOren ni

First of all, the NEG marker is /ni/ instead of the expected /na/. Second, the perfect aspect markings (/ye/ followed by /ch/) are missing. Third, the marking of Present vs. Past is neutralized. Only the Present form is allowed, regardless of the intended meaning (which may be either past or present). Satisfying historical explanations for the development in Bangla of this peculiar method of Perfect negation have yet to emerge. Nor is it clear what devices to use in a synchronic description of the phenomenon, which is moderately difficult for non-native learners to pick up.

Suppose that we were to place the descriptive burden on the transformational component. A transformation called Negative Perfect Neutralization, which I state as (82), would do the job, at a point in the transformational derivation where NEG had reached its eventual position. The point of ordering (82) after NEG-moving transformations would be to preclude misderivations such as (83i-iii).

(82) Negative Perfect Neutralization (obligatory)
S.D: V – PAST – /ch/ – M – T – PERS – NEG
S.C: 1 2 3 4 5 6 7 ➔1, 0, 0, 0, 5, 6, 7+2

(83) i. $[_{PP}[_{S}$ uni eTa korechen na$][_{P}$bole]] ami biSSaS kortam
"(s) he-HON this-item has-done not as I belief would-do"
'I'd believe that (s)he hadn't done this'
ii. uni eTa kOren ni bole ami biSSaS kortam
"(s)he-HON this-item does not-PERF as I belief would-do"
iii. *uni eTa kOren bole ami biSSaS kortam ni
"(s)he-HON this-item does as I belief would-do not-PERF"

The misderivation in (83) works like this. Application of (82) to (83-i) yields (83-ii). If the optional rule of NEG-Transportation, which I have not stated, were to apply to (83-ii), the result would be (83-iii), which is ungrammatical. But we have ordered (82) after all NEG-moving transformations, including NEG-Transportation. So (82-ii) will never be available as input to NEG-Transportation, and the misgeneration of (83-iii) will not take place.

One problem with (82) is that it is too 'complicated': it requires a stipulation of ordering and an S.D. which is unusually complex for a movement rule. This problem is not too serious. We can reformulate (82) as a deletion-and-copying rule. That move places it in the component of Deletion Rules. In several current frameworks (e.g. Chomsky (1978)), the component of Deletion Rules as a whole is held to follow that of Movement Rules. So it need not be stipulated that (82) follows rules that move NEG. And the S.D. of (82) is not unusually complex for a Deletion Rule. In order for the operation to be copying instead of movement, the S.C. will have to say $\begin{bmatrix} 7 \\ 2 \end{bmatrix}$ where it now says 7+2; the output, instead of reading $[_{NEG}$NEG

PAST] for that term, will now read $\begin{bmatrix} \text{NEG} \\ \text{PAST} \end{bmatrix}$.

A more serious problem with (82), which changing it into a Deletion Rule wouldn't solve, is that (82) violates the condition of recoverability. It irrecoverably deletes the fourth factor of the operand, the factor which carries the distinction between PRESENT and PAST mood. Given the fundamental importance of recoverability of deletion in generative grammar, it seems necessary to reject (82) and, accordingly, to shift the burden of describing perfect negatives to the component of Inflection Rules. I propose the following additions to the framework of chapter 4. Surely the next investigator of this subject will find sound reasons for rejecting (86). I am glad that nothing crucial in the present volume rests on my analysis of perfect negatives.

(84) S.D.: $[_V (V) - V]$
S.D.: 1 2 ➔ 1, 2+ $[_{ASP}$ ±Perfect] (optional)

(85) S.D.: $[_V (V) - [_V V$ ASP]]
S.C.: 1 2 ➔ 1, 2+V

(86) S.D.: $[_S X$ – +Perf – /ch/ – M – Future – PERS (HON) – NEG]
S.C.: 1 2 3 4 5 6 7
➔1, 0, 0, Present, 5, 6, 2

(84)-(85) precede (IV-66)-(IV-68). (86) follows (IV-73). The formulation of (86) is based on the assumption that, since interpreting the instruction to replace 7 by 2 as an instruction to remove the node NEG and its contents and put term 2 in its place would infringe the principle of recoverability of deletion, the instruction to replace 7 by 2 – in deference to that principle – is actually interpreted as an instruction to make factor 2 of the operand a feature on the NEG node in factor 7. If, as has been assumed by many since Chomsky (1970), every lexical category amounts to a bundle of categorial features, then the result of adding +Perfect to NEG will simply be a slightly enriched feature bundle. However, I am not aware of actual research on a feature analysis of the node NEG. There is no way, currently, to decompose the seventh term in the output of (86) into 'universal features' of the sort explored in Jackendoff (1977).

The formulation of (84) ensures that the pole of a compound verb will never be aspectually marked. This is in fact true. The formulation also correctly predicts that the head V of an infinitive phrase may be aspectually marked. There is, however, a constraint, not expressed by (84), against an infinitive phrase's head V being in the Progressive aspect, since a repetition of the 'Infinitive' /yte/ suffix (shared by the Progressive aspect and by the untensed Present mood) is not tolerated. Presumably a surface constraint against /yte/-repetition takes care of this. Such anti-repetition surface constraints are common. It is unfortunate that no one has yet gone to the trouble of surveying such constraints with a view to determining their distribution in language families, whether they prohibit repetition only of affixes, or of some specifiable subset thereof, etc.

[*Clarification added in 2020: I should have specified how I expected rule (85) to operate. I would like to clarify now, belatedly, that I assumed that the bald 'V' in the S.C. of (85) would trigger the insertion of some default V, to wit, the copula shown as /ch/ in (86), and would pave the way for the formation of an aspect-marked finite verbal word.*]

Since the change of term 4 in rule (86) is not a deletion, it does not count as infringing the recoverability condition. As for term 3, /ch/ may be regarded as a designated element of the sort that may be freely deleted. Thus, the relocation from the syntactic to the inflectional component of the grammar has been a useful move.

Rule (86) mentions /ch/ instead of merely mentioning V. The auxiliary V slot is filled by /ch/ only when the Mood of the verb is non-Imperative and the Tense is non-Future. Otherwise, the auxiliary used is /thak/. See section 2.3.2 for the pattern. I should add here that there is a main verb use of /ch/. In this use, /ch/ appears as /ach/ in the Present Mood and /ch/ in the Past Mood. In this use as well as in the auxiliary use, /ch/ is defective, in that it occurs only when the Mood is non-Imperative and the Tense non-Future. The two uses of /ch/ are no doubt connected. Nor do I doubt that the way in which the perfect negative marker /ni/ 'replaces' /ye-ch-..na/ is somehow connected with the fact that, instead of the expected negative forms /ache na/, /achi na/ etc. in the Simple Present of the main verb /ach/, we get the invariable Simple Present negative form /ney/ which shows no change for Person or Honour, rather like substandard English *ain't*. But I offer no description of these connections here.

5.5 Consequences of non-clausality

In the last few sections I have been dealing with a variety of grammatical facts and showing how the different regions of my analysis of Bangla work together. Even where I was not giving specific arguments for one contention or another, there was always the tacit claim that the various analyses made were upheld by my more controversial proposals, the ones which I am particularly proposing as 'cornerstones' of my grammar. One of these controversial proposals is the hypothesis, formulated and defended in section 5.1, that infinitive structures, whatever they may be in language such as English, are not clauses in Bangla. In sections 5.2-5.4 I have presented analyses within my framework of Bangla grammar which, more or less indirectly, support the claim that Bangla infinitive structures are not clauses. I now return to a more direct defence of this claim. I will show that this claim implies a controversial prediction about an area of grammar where the claim I am attacking (CIH, the Clausal Infinitive Hypothesis) makes the opposite prediction, and that the prediction which follows from my claim is true while the other claim leads to a false prediction.

If, in Bangla, infinitive structures were clauses, they would have had the main properties of bona fide clauses. They would, for instance, be able to be relative clauses and constituent questions. The prediction that infinitive structures should have these abilities follows from CIH. My proposal, NCIH, predicts instead that infinitive structures, unlike bona fide clauses, cannot be relative clauses or constituent questions. For the time being, let us make and test the opposed predictions without trying to understand quite what is involved. We know what we need to know – that there is, for some reason, a correlation between being an S and being able to be a relative clause or a constituent question.

The data shows that the view that infinitives in Bangla are not clauses is correct and the contrary view incorrect. (For some readers it may be necessary to specify that the strings given in (87)-(90), ill-formed if /te/ is taken to be the infinitive marker, are irrelevantly well-formed if /te/ is construed as the past habitual third person non-honorific ending instead.)

(87) *jake jigeS korte e Sey lok

“whom question do-INF this that person”
‘This is the person (whom) to ask’

(88) *je churi diye hire kaTte e Sey churi
“which knife with diamond cut-INF this that knife"
‘This is the knife with which to cut diamonds’

(89) *kake jigeS korte ta jani na
"whom? question do-INF that (I-)know not”
‘I don’t know whom to ask’

(90) *kon churi diye hire kaTte ta jani na
“which? knife with diamond cut-INF that (I-)know not”
‘I don’t know which knife to cut diamonds with’

(91) jake jigeS korte caW e Sey lok
“whom question do-INF (you-)want this that person”
‘This is the person whom you want to ask’

(92) jake jigeS korbe e Sey lok
"whom question (you-)will-do this that person”
‘This is the person whom you will ask’

(93) je churi diye hire kaTte parbe e Sey churi
“which knife with diamond cut-INF (you-)will-be-able this that knife”
‘This is the knife with which you’ll be able to cut diamonds’

(94) je churi diye hire kaTbe e Sey churi
“which knife with diamond (you-)will-cut this that knife”
‘This is the knife with which you’ll cut diamonds’

(95) kake jigeS korte caW ta jani na
“whom? question do-INF (you-)want that (I-)know not”
‘I don’t know whom you want to ask’

(96) kake jigeS korbe ta jani na
"whom? question (you-)will-do that (I-)know not”
‘I don’t know whom you will ask’

(97) kon churi diye hire kaTte parbe ta jani na
“which? knife with diamond cut-INF (you-)will-be-able that (I-)know not”
‘I don’t know which knife you’ll be able to cut the diamonds with’

(98) kon churi diye hire kaTbe ta jani na
“which? knife with diamond (you-)will-cut that (I-)know not”
‘I don't know which knife you’ll cut diamonds with’

The hypothesis that infinitive structures are clauselike entails that (87)-(90) should be well-formed just as (91)-(98) are. But this is not so. Therefore, infinitive structures are not clauselike. Now we ask what properties of clauses were involved in this contest between the hypotheses. We ask what it is about clauses that enables them to be relative clauses as in (91)-(94) or constituent questions in (95)-(98).

There is no currently available sense of 'semantic' in which this question can be said to have a semantic answer. Whatever one's favourite semantic theory may be, all current theories of semantics heavily rely on the possibility of paraphrase and translation. The application of a term like Infinitive to constructions in a variety of language in practice relies on the possibility of translating Infinitives of one language into infinitives of other languages. As far as translation is concerned, Bangla infinitives seem more or less equivalent to English infinitives, although there are discrepancies in the many positions where, as discussed in chapter 4, gerund phrases block infinitive phrases in Bangla and thus gerunds have to step in to translate English Infinitives. But English infinitives, as the glosses indicating the intended interpretations of (87)-(90) show, do exhibit the clausal property of being able to be relativized or constituent-questioned. So, it is not a matter of the 'semantic' properties which translation can tap. The question 'What is it about a clause that enables it to be a relative clause or a constituent question?' thus must be construed as a syntactic question, as a question about the structure which the various parts of Bangla syntax must assign to Bangla clauses so as to enable these clauses, without at the same time enabling Infinitive structures, to be relativized or constituent-questioned.

Clearly, the visible thing about a relative clause or a constituent question that makes it what it is is the fact that it contains an NP or AP whose determiner is a relative word in *j*- or an interrogative word in *k*-. But that clearly is not the criterion. The infinitive phrases in the ill-formed (87)-(90) contain such *j*- and *k*-phrases, and yet they do not count as relative or interrogative clauses. Also, the crucial thing cannot be that the S node of a relative clause or constituent question should directly dominate the NP or AP node whose determiner is a relative or interrogative word. In the well-formed (91), (93), (95), and (97) it is the infinitival VP node which directly dominates the sensitive NP node, and yet in these cases the S node immediately dominating this VP node is a relative or interrogative clause.

Such exercises show that taxonomic methods are not going to work. The analyst who wants to study Bangla relative or interrogative structures must come up with a solution based partly on data and partly on conjecture (or imagination, if one prefers that term), make predictions based on this solution, and test the predictions against further data.

5.6 A theory of clause structure

I conjecture that a node dominating a relative or interrogative clause is not S, but S′. Departing from the style of current proposals for the transformational grammar of other languages, I refrain from making the further claim that S′ rather than S is the only initial symbol of the PS subcomponent. Rather, resuscitating a never much discussed proposal by Chomsky (1957) that there be several initial symbols for the PS rules, I propose that at least S and S′ be such initial symbols. The present study will not undertake further expansion of the list of initial symbols, although such expansion may eventually prove necessary. My rewrite for S′ is S COMP, for reasons that will become clear in chapter 6. If a clause is relative, COMP dominates that feature

[–INTERROGATIVE], abbreviated as [–INT]. If a clause is interrogative, COMP dominates +INT. One may ask whether it would make any difference if ±INT were a feature on COMP instead of under COMP. I assume, arbitrarily, that the answer is no. The question needs to be pursued in theoretical research.

(99) i. S′ → S COMP
ii. COMP → ±INT

If a clause is neither relative nor interrogative, it is an S and not an S′, we will assume at this point. (This assumption will change in chapter 6.)

Let us provisionally assume some mechanism for matching up the ±INT marking on the COMP of an S′ with the ±INT markings on the DET nodes in the clause. A well-formed relative S′ will have (i) at least one DET node dominating a –INT word, (ii) no DET dominating a +INT word, and (iii) a COMP which bears the feature –INT. To get the corresponding statement for a well-formed constituent question, read + for – and – for +.

The discussion of NP structure in section 3.1 showed, among other things, that relative pronouns beginning with *j*- and interrogative pronouns beginning with *k*- are actually sequences of a relative or interrogative determiner plus either the Honorific Noun /ni/ or a Zero Noun. Let us analyze these determiners, the basic ingredient of the clause types we are concerned with here, as follows. If a DET is either relative or interrogative, its lexical entry has a feature +OPEN. A relative DET is +OPEN –INT; an interrogative DET is +OPEN +INT. In this chapter it is being assumed that only relative and interrogative clauses are dominated by the node S′. So, we can interpret +OPEN as an instruction to insert DET in an S immediately dominated by S′, or, equivalently, in an S with a COMP sister. This is only a preliminary definition which becomes invalid in chapter 6, for the range of S′ will be enlarged. The notion +OPEN will then need redefinition. To summarize:

(100) Determiners are lexically specified as +OPEN. +OPEN Determiner entries are further specified as ±INTERROGATIVE or ±INT. +OPEN +INT Determiners begin with a /k/. [+OPEN, –INT] or Relative Determiners begin with a /j/. For well-formedness, the value of the INT feature for all INT-specified Determiners in a clause must be the same as the value of INT for that clause's Complementizer.

The reason why infinitive phrases cannot be relative clauses or constituent questions, then, is that the infinitive structure, being a Verb Phrase and not an S′, has no COMP to 'take responsibility for' +OPEN Determiners.

My attempt to explain the difference between infinitives and clauses has invoked an entity – the COMP node – which has neither been seen nor heard in the structures examined so far. What predictions follow from the postulation of a COMP node? Only if the predictions, whatever they are, are correct is the postulation of this node permissible. Accordingly, chapter 6 will make and test predictions based on rule (99-i) which introduces COMP and on rules which presuppose (99-i). Facts corroborating (99-i) will justify the use of the Complementizer concept.

Chapter 6

COMPLEMENTIZERS, CONJUNCTIONS, EMPHASIZERS

6.0 Strategy and tactics

Chapter 5 sought to explain some differences between clauses and infinitive phrases by postulating that a clause, an S′, contains an extra node, the COMP node generated by the PS rule S′ → S COMP, which the infinitive lacks, being a VP. This PS rule needs justification and the grammar of complementizers needs to be worked out in relation to (i) the class of conjunctions with which the class of complementizers overlaps and (ii) the class of emphasizers with which complementizers enter into special sequential relations when they co-occur.

In this chapter, the first task is to provide 'primary' justification for the rule expanding S′ – data confirming the direct observational predictions that follow from various aspects of the rule.

This done, any success in attempts to base explanations of further phenomena on the tentatively confirmed theory of clause structure will constitute 'secondary' justification for the decision to posit a COMP node. On the basis of sentences where the COMP occurs in various positions, a movement rule can be postulated which preposes COMP into S. The proper application of this rule requires a principle banning all movement rule applications which break into sequences consisting of head and modifiers. This principle, although this chapter states it conservatively with a view to carrying out just the task at hand, follows from Mark Baltin's (1978) theory of movement rules.

I will consider next the problems raised by those words which function both as complementizers and as conjunctions. Evidence will be given that a theory which distinguishes the two classes and regards some complementizers as homonymous to some conjunctions is better able to handle the facts than a theory, offered by J. Chattopadhyay (1976a, b), which regards the complementizer /je/ and the conjunction /je/ as the same element.

I should, perhaps, rephrase that. J. Chattopadhyay postulates a single element /je/ which she regards just as a complementizer. What I object to is that she does not see /je/ as sometimes functioning as a conjunction. In my theory, there are some clear cases of conjunctions which are never complementizers (e.g. *ebong* 'and') and of complementizers which are never conjunctions (e.g. *bujhi* 'whether'). In addition, there are words like /je/ 'that' which occur in both roles – analyzable either as a single element which is both CONJ and COMP in its lexical entry or as two distinct lexical entries (one CONJ and one COMP) which are homonymous, although, if one opts for the latter analysis, one should set up something systematic, like a redundancy rule, to express the fact that a coherent set of CONJ words bears a one-to-one homonymy relation to a coherent set of COMP words.

Turning from relations of similarity to those of contiguity, I will finally examine emphasizers and their relation to complementizers. The study of emphasizers is a necessary preliminary to the study of indefinite expressions in subsequent chapters, since an existing lexicographic analysis claims – incorrectly, to my mind – that indefinite expressions contain an emphasizer.

6.1 Complementizers

If a COMP node exists, then at least in some trees it must dominate a tangible morph or two, for otherwise Occam's razor will scrap it. The first prediction of (V-99) which needs testing, then, is the prediction that some clauses possess a visible complementizer. The prediction is correct:

(1) $[_{S'}[_{S}[_{NP}$ tomar$][_{NP}$ doy$][_{V}[_{A}$ bhal$][_{V}$ lage$]]][_{COMP}$ bujhi$]]$
"you-GEN yogurt good feels whether"
'Do you like yogurt?'

(2) $[_{S'}[_{S}[_{NP}$ tomar$][_{NP}$ doy$][_{V}[_{A}$bhal$][_{V}$ lage$]]][_{COMP}$ kintu$]]$
"you-GEN yogurt good feels, however"
'You like yogurt, however'

The second part of (V-99), the rule rewriting COMP as ±INT, in conjunction with the tacit assumption in chapter 5 that a COMP must be either +INT or –INT, predicts the existence of two kinds of complementizer words: those which enter $[_{COMP}\ [_{-INT}\ \_\_\_\_]]$, and those which enter $[_{COMP}\ [_{+INT}\ \_\_\_\_\_]]$. One also expects that these two subclasses of COMP words will include at least one member each which are respectively /j/-initial and –INT, and /k/-initial and +INT. It is not that the presence of other members in these classes (or the existence of a class of COMPs whose members are neither + nor –INT) will controvert (V-99), but rather that the existence of a /j/-initial –INT complementizer and of a /k/-initial +INT complementizer will strongly corroborate (V-99) and (V-100).

This prediction is confirmed too. The words *bujhi* in (1) and *ki* in (3), which respectively don't and do begin with the hoped-for /k/, enter a $[_{COMP}\ [_{+INT}\_\_]]$ slot. The slot for *je* in (4) – and the word does begin with the expected /j/ – is $[_{COMP}\ [_{-INT}\_\_]]$. In addition, one finds that one needs to treat the complementizer *kintu* of (2) as unspecified for the feature INT. This remark must remain obscure and intuition-bound until Interrogativity has been defined with more clarity. I will return to this matter in chapter 7.

(3) $[_{S'}\ [_{S}\ [_{NP}$ tomar$][_{NP}$ doy$][_{V}[_{A}$ bhal$][_{V}$ lage$]]][_{COMP}$ ki$]]$
"you-GEN yogurt good feels whether"
'Do you like yogurt?'

(4) $[_{S'}\ [_{S}\ [_{NP}$ tomar$][_{NP}$ doy$][_{V}[_{A}$ bhal$][_{V}$ lage$]]][_{COMP}$ je$]]$
"you-GEN yogurt good feels that"
'It's just that you like yogurt'

Another direct prediction made by rule (V-99-i) is that a clause is syntactically unentitled to more than one complementizer. One way to test this prediction is by constructing two-COMP sentences which, as far as one can judge such matters, should be semantically interpretable. It turns out that sentences of this sort, like (5)-(6), are what it has become usual to call 'crashingly ungrammatical'. Corroboration again.

(5) *tomar doy bhal lage *ki kintu*

"you-GEN yogurt good feels *whether however*"
'Do you like yogurt, however?'

(6) *tomar doy bhal lage *je bujhi*
"you-GEN yogurt good feels *that whether*"
'Is it just that you like yogurt?'

Finally, we need to test the claim made by rule (V-99-i) that S is a single constituent whose sister is COMP. How do we know? The answer lies in coordinated sentences like (7) and (8).

(7) $[_{S'}[_{S}$ bap ma mara gEche$][_{CONJ}$ ba$][_{S}$ ghOr dor bheSe gEche$][_{COMP}$ bujhi$]]$
"father mother dead have-gone or house door floating have-gone whether"
'Have their parents died or their property been flooded? (has one of these things happened?)'

(8) $[_{S'}[_{S}$ bhikkhe$[_{EMP}$ o$]$ nebe na$][_{S}$ biddroho$[_{EMP}$ o$]$korbe na$][_{COMP}$ je$]]$
"alms-too will-take not rebellion-too will-do not that"
'that they'll neither take alms nor rebel'

Sentences like (7)-(8), where the structure shown abbreviates what I take to be the correct structure – $[_{S'}[_{S}$ S (CONJ) S][COMP]] – cannot be generated if one assumes either that COMP is a daughter of S or that COMP is a sister-daughter of S. This is obvious where COMP is a daughter of S, less obvious in the latter case. A demonstration for the second case would have to be deferred until the more detailed look at coordinate structures later in the chapter. Luckily, it is unnecessary to depend on an argument that will only be provided at a later juncture. There is another reason for rejecting $[_{S}$ S COMP], one that can be mentioned at once. Namely, such an analysis would fail to explain why there is only one complementizer per clause, since it would allow structures like $[_{S}[_{S}[_{S}$ S COMP] COMP] COMP]. So, we can take it as established that the rules in (V-99) are correct with respect to the direct tests one can carry out. The only thing we need to modify is the idea that (V-99-ii) might be in effect obligatory. We need to assume, instead, that it is optional, yielding three sorts of complementizers (+INT, –INT, neither) and not just two.

(V-99) i. S′ → S COMP
ii. COMP → ±INT

6.2 Complementizer Preposing

In the following sentences, the COMP occurs in non-final position.

(9) tomar doy *bujhi* bhal lage "you-GEN yogurt whether good feels"

(10) tomar *bujhi* doy bhal lage "you-GEN whether yogurt good feels"

(11) **bujhi* tomar doy bhal lage "whether you-GEN yogurt good feels"

(12) tomar doy *kintu* bhal lage "you-GEN yogurt however good feels"

(13) tomar *kintu* doy bhal lage "you-GEN however yogurt good feels"

(14) *kintu* tomar doy bhal lage "but you-GEN yogurt good feels"

(15) tomar doy *ki* bhal lage "you-GEN yogurt whether good feels"

(16) tomar *ki* doy bhal lage "you-GEN whether yogurt good feels"

(17) **ki* tomar doy bhal lage "whether you-GEN yogurt good feels"

(18) tomar doy *je* bhal lage "you-GEN yogurt that good feels"

(19) tomar *je* doy bhal lage "you-GEN that yogurt good feels"

(20) **je* tomar doy bhal lage "that you-GEN yogurt good feels"

The pattern, disturbed by the exceptional (14), is that the COMP in an S′ may occur between any two immediate constituents of the S in addition to being able to follow the S as a whole, but cannot occur at the beginning of the clause. Before attempting a formalization of the pattern let us examine (14) a bit more closely. In all the other cases of COMP movement, a sentence with some complementizer C_j has the same basic meaning regardless of the position of C_j in it (the focus of the sentence depends on the position of C_j in ways which cannot be accurately shown in one-sentence English renderings and which will, for reasons of space and relative unimportance, be ignored in this volume); but (14) has a special gloss. To the extent that English 'You like yogurt, however' differs from 'But you like yogurt', the latter is the meaning of (14) and the former the meaning of (2), (12), and (13). So, we suspect that *kintu* of (14) is not the complementizer *kintu* but a semantically related homonym belonging to a different grammatical category, somewhat like the adjectival and prepositional versions of the English word *near*. What confirms this suspicion is the fact that the *kintu* of (14) permits a complementizer clause-mate, as shown in (21), whereas we see in (22)-(24) that adding such a clause-mate becomes impossible when *kintu* occurs in one of the complementizer positions:

(21) [$_U$ kintu] tomar doy bhal lage [$_{COMP}$ je]
"but you-GEN yogurt good feels that"
'But it's just that you like yogurt'

(22) *tomar doy [$_{COMP}$ kintu] bhal lage [$_{COMP}$ je]
"you-GEN yogurt however good feels that"

(23) *tomar [$_{COMP}$ kintu] doy bhal lage [$_{COMP}$ je]
"you-GEN however yogurt good feels that"

(24) *tomar doy bhal lage [$_{COMP}$ kintu][$_{COMP}$ je]

"you-GEN yogurt good feels however that"

The well-formedness of (21) in contrast to the ill-formedness of (22)-(24) indicates that the U(nknown) node in it ≠ COMP. Maybe U = CONJ(unction). So (14), with its *kintu* regarded as a complementizer, is simply not to be generated. Having thus disposed of the exception, let us try and formalize the rule responsible for the facts shown at (9)-(20).

(25) Complementizer Preposing
S.D.: X – X″ – Y – COMP
S.C.: 1 2 3 4 ➔ 1, 2+4, 3, 0

Not directly useful for advancing our inquiry but nevertheless pertinent is the observation that if (25) formulates the rule correctly then one must construe the various absences in clauses – e.g. the absence of a subject in *cole jacchi je* "leave-and am/are-going that" 'It's just that I'm (or we're) going away' – as the relevant node really going missing, not as a node which secretly sits there dominating something (a zero noun) or nothing (*e*). For consider what would happen if *cole jacchi je* were in fact to be represented as $[_{S'}[_{S}\,[_{NP}\,e]$ cole jacchi$][_{COMP}$ je$]]$. Rule (25) would then apply, to right-adjoin *je* to the empty NP. The output *Ø je cole jacchi* would be generated, with the complementizer in effect clause-initial since nothing audibly precedes it; but this output is, in point of fact, ill-formed. We conclude that *cole jacchi je* does lack an NP node at the point at which COMP Preposing applies.

Suppose, for a moment, that sentences that appear without an overt subject are base-generated with a subject but then lose it by a deletion rule which, being a deletion rule, applies after all movement rules have applied. On this hypothesis, the deep structure of *cole jacchi je* would have either *ami cole jacchi je* or *amra cole jacchi je* as its terminal string (the choice is between *ami* 'I' and *amra* 'we', which trigger identical agreement on the verb in Bangla). Let us suppose it is *ami cole jacchi je* 'It's just that I'm going away'. Now, since COMP Preposing is, like other movement rules, optional, there will be two possible outcomes, *ami cole jacchi je* (no preposing) and $[_{NP}\,[_{NP}\,ami][_{COMP}\,je]]$ *cole jacchi* (preposing has applied using the standard procedure of adjunction). In the first case, when the time comes to decide whether to apply the optional rule deleting a minimally specific subject, *ami* can be routinely deleted, since the information it contains remains encoded in the verb suffix (except that the singular number is lost, which Bangla deletion rules, for some reason, seem not to care about). In the case of trying to apply the rule of subject deletion to $[_{NP}\,[_{NP}\,ami][je]]$ *cole jacchi*, on the other hand, first of all the deletion rule cannot touch the lower NP *ami*, because of A-over-A, and secondly it cannot touch the entire NP *ami je* because the information contained in the adjoined complementizer *je* would be irretrievable if lost. So deletion does not apply in this case.

The above discussion indicates that it is useful that (25) effects adjunction rather than sister-adjunction, for in the latter case it would have been possible for the deletion rule to remove *ami* from *ami je cole jacchi*, leaving the ill-formed output **je cole jacchi*.

However, since it is not apparent that there should be a rule of subject deletion (an obvious alternative being to make the first NP node of the S expansion optional), the above consideration does not suffice to establish the desirability of, say, (26-a) rather than (26-b) for (18).

(26) a. $[_{S'}[_{S}\,[_{NP}$ tomar$][_{NP}[_{NP}$ doy$][_{COMP}$ je$]][_{V}[_{A}$ bhal$][_{V}$ lage$]]]]$

b. $[_{S'}[_{S}\,[_{NP}$ tomar$][_{NP}$ doy$][_{COMP}$ je$][_{V}[_{A}$ bhal$][_{V}$ lage$]]]]$

One might deny the need to defend (26-a). Evidence from other languages, one might argue, already sufficiently motivates the theory of adjunction which engenders (26-a). There is, moreover, no evidence supporting (26-b) against (26-a). The burden of proof does rest on a proponent of (26-b). However, since I have found two facts which positively support (26-a) against (26-b), I feel that I should mention them.

The first fact is that, of the two forms of the conjunct verb stem $[_{V}[_{A}$ bhalo$][_{V}$ lag$]]$ – the colloquial or fast form *bhal lag* and the careful or slow form *bhalo lag* – only the slow form can be broken into by a preposed complementizer, and that, likewise, of the two forms of the compound verb stem $[_{V}[_{V}$ boSe$][_{V}$ thak$]]$ – the fast form *boS thak* and the slow form *boSe thak* – only the slow form can be broken into by a preposed complementizer. Examples follow.

(27) $[_{S'}[_{S}\,[_{NP}$ tomar$][_{NP}$ doy$][_{V}[_{A}$ bhalo$][_{V}$ lage$]]][_{COMP}$ je$]]$

(28) $[_{S'}[_{S}\,[_{NP}$ tomar$][_{NP}$ doy$][_{V}[_{A}[_{A}$ bhalo$][_{COMP}$ je$]][_{V}$ lage$]]]]$

(29) $*[_{S'}[_{S}\,[_{NP}$ tomar$][_{NP}$ doy$][_{V}[_{A}[_{A}$ bhal$][_{COMP}$ je$]][_{V}$ lage$]]]]$

(30) $[_{S'}[_{S}\,[_{NP}$ ora$][_{V}[_{V}$ boSe$][_{V}$ thakbe$]]][_{COMP}$ ki$]]$
"they sit-and will stay whether"
'Will they keep waiting?'

(31) $[_{S'}[_{S}\,[_{NP}$ ora$][_{V}[_{V}[_{V}$ boSe$][_{COMP}$ ki$]][_{V}$ thakbe$]]]]$

(32) $[_{S'}[_{S}\,[_{NP}$ ora$][_{V}[_{V}$ boS$][_{V}$ thakbe$]]][_{COMP}$ ki$]]$

(33) $*[_{S'}[_{S}\,[_{NP}$ ora$][_{V}[_{V}[_{V}$ boS$][_{COMP}$ ki$]][_{V}$ thakbe$]]]]$

It seems plausible that, since /bhalo/ 'good' and /boSe/ 'sitting' never have the phonetic forms *bhal* and *boS* except when these two-word verb stems are used in fast speech, they should be regarded as proclitics when they have these special contracted forms. It follows that they cannot themselves take clitics, since in general clitics do not themselves take clitics. Thus, one would have an explanation for the ill-formedness of (29) and (33) if one were to assume that a preposed complementizer /je/ is an enclitic to the constituent to which it is adjoined. Since there seems to be no other explanation for the ill-formedness of (29) and (33), it thus becomes desirable to assume that the preposed complementizer is encliticized – to assume, then, that the structure of (18) is (26-a) and not (26-b).

The second piece of evidence comes from orthography, which, as Sapir noted, sometimes reflects the grammatical intuition of native speakers. Many preposed complementizers are written with a hyphen joining the complementizer to its left neighbour. Some preposed complementizers are written with neither space nor a hyphen separating the complementizer from its left neighbour (written Bangla uses spaces, as written English does, to demarcate words). Such use of hyphenation or direct junction is especially common where the complementizer is formally identical to a determiner and makes the clause ambiguous (for example, *tomar je doy bhal lage* has the $[_{DET}$ je] reading "you-GEN which yogurt good feels"

'the yogurt which you like' and the [$_{COMP}$ je] reading "you-GEN that yogurt good feels" 'It's just that you like yogurt'). Although this fact by itself would not prove that (26-a) is a better structure for (18) than (26-b), the fact seems to me to align with the first fact and with the independent support for the theory of adjunction. So, this bit of evidence, though it would get us nowhere on its own, does reinforce the case for (26-a).

Notice, though, that rule (25) as it stands will not handle the sentences (27)-(33) which we have used as data. (25) will adjoin a complementizer only to an X″, and the A of (27) and the V of (31) are obviously not X″ nodes on any analysis. One reformulation for (25) would be (34), which does not specify that Y must be a single constituent but which the adjunction convention will permit to apply only where it factors the phrase-marker in such a way as to match Y up with a single constituent, since otherwise the adjunction routine is unimplementable.

(34) Complementizer Preposing restated

S.D. : X – Y – Z – COMP
S.C. : 1 2 3 4 ➔ 1, 2+4, 3, 0

The problem with (34) is that it wrongly predicts that a sentence like (35), which is (21) minus a few specifications of structure, will transform to (36).

(35) [$_{S'}$ [$_{U}$ kintu][$_{NP}$ tomar][$_{NP}$ doy][$_{V}$[$_{A}$ bhal][$_{V}$ lage]][$_{COMP}$ je]]

(36) *[$_{S'}$ [$_{U}$ [$_{U}$ kintu][$_{COMP}$ je]][$_{NP}$ tomar][$_{NP}$ doy][$_{V}$[$_{A}$ bhal][$_{V}$ lage]]]

We decided, remember, that U(nknown) may well be CONJ, but is not COMP. How are we to block (36) without blocking (31)? Some options:

(37) Option one: Restate the S.D. of (34) as X – L – Z – COMP, where L is a variable ranging over all lexical and major nodes. Conjunctions are neither.

Option two: Put *kintu* in (35) outside S but inside S′ and restate the S.D. of (34) as [$_{S}$ X – Y – Z] – COMP.

Option three: Claim that *kintu* is a proclitic and therefore, like *bhal* in (29), is not entitled to take an enclitic.

Option four: Put *kintu* in (35) outside S but inside S′, leave (34) as it stands, and devise a principle that will prevent structure-destroying movement rules from inserting material into either of the positions marked _ in [$_{PH}$ …_...H…_...], where H is the head of PH.

There is evidence against options one, two, and three. Option two wrongly predicts that (38) should transform to (39) and options one to three all wrongly predict that (38) should transform

to (40).[8] Option four remains as the only serious contender.

(38) $[_{S'}[_{S}\,[_{VP}\,[_{NP}$ oSodhi$][_{V}[_{NEG}$ na$][_{V}$ ante$]]][_{NP}$ honumanke$][_{V}$ bolecho$]][_{COMP}$ ki$]]$ "herb not bring Hanuman-OBJ have-told whether"
'Have you told Hanuman not to bring the herbs?'

(39) $*[_{S'}[_{S}[_{VP}[_{NP}$ oSodhi$][_{V}[_{NEG}[_{NEG}$ na$][_{COMP}$ ki$]][_{V}$ ante$]]][_{NP}$ honumanke$][_{V}$ bolecho$]]]$
"herb not whether bring Hanuman-OBJ have-told"

(40) $*[_{S'}[_{S}\,[_{VP}[_{NP}[_{NP}$ oSodhi$][_{COMP}$ ki$]][_{V}[_{NEG}$ na$][_{V}$ ante$]]][_{NP}$ honumanke$]$
$[_{V}$ bolecho$]]]$
"herb whether not bring Hanuman-OBJ have-told"

(41) $[_{S'}[_{S}[_{NP}$ honumanke$][_{NP}$ oSodhi$][_{V}[_{V}[_{NEG}$ na$][_{V}$ ante$]][_{V}$ bolecho$]]\,[_{COMP}$ ki$]]$
"Hanuman-OBJ herb not bring have-told whether"
'Have you told Hanuman not to bring the herbs?'

(42) $*[_{S'}[_{S}[_{NP}$ honumanke$][_{NP}$ oSodhi$][_{V}[_{V}[_{NEG}[_{NEG}$ na$][_{COMP}$ ki$]][_{V}$ ante$]]$
$[_{V}$ bolecho$]]]$
"Hanuman-OBJ herb not whether bring have-told"

The only problem with option four is that it does not preclude the derivation of (42) from (41). But this is a soluble problem. It is possible simply to place the burden of handling this problem on the theory of adjunction. Just as it is the routine of adjunction which in practice guarantees that Y in the S.D. of (34) will always correspond to a single node in a well-formed derivation, so also we can assume that only members of the X-bar system ever get things adjoined to them, and that therefore (34) will never analyze a phrase-marker in such a way that Y matches up with a node like CONJ or NEG which is outside the X-bar system. Thus, we now have two reasons for the ungrammaticality of (36) and (39) – that they infringe, first, the principle posited in option four of (37), and second, the conventions of adjunction. In contrast, (40) is ill-formed only for the first and (42) only for the second reason.

Another advantage of positing the principle mentioned in option four, a principle which we may state as (43), is that it correctly predicts that (44) should fail to transform to (45) or (46).

(43) **The No New Satellite Condition**: Given $[_{H'}$ X_Y H Z_W$]$, where H is a projection of a bar-notational feature bundle and is the head of H′ and where X, Y, Z, W (if non-null) are satellites of H, no rule may move material into either of the positions marked _.

(44) $[_{S'}[_{S}[_{NP}[_{AP}[_{NP}$ hemontor$][_{A}[_{V}$ lekh$]$ a$]][_{N}$ boy$]][_{V}[_{V}$ poRben$][_{NEG}$ na$]]][_{COMP}$ bujhi$]]$
"Hemonto-GEN written book will-read not whether"
'Won't you read the book written by Hemonto?'

[8] Footnote added in 2020: The **string** in (40) is in fact acceptable, with intonation suggesting a structure that places *oSodhi* in a topic position. We need not revisit the conclusion that **structure** (40) is ill-formed.

(45) *[$_{S'}$[$_{S}$[$_{NP}$[$_{AP}$[$_{NP}$[$_{NP}$ hemontor][$_{COMP}$ bujhi]][$_{A}$[$_{V}$ lekh] a]][$_{N}$ boy]][$_{V}$[$_{V}$ poRben] na]]]
"Hemonto-GEN whether written book will read not"

(46) i. *[$_{S'}$[$_{S}$[$_{NP}$[$_{AP}$[$_{AP}$[$_{NP}$ hemontor][$_{A}$[$_{V}$ lekh] a]][$_{COMP}$ bujhi]][$_{N}$ boy]][$_{V}$[$_{V}$ poRben] na]]
"Hemonto-GEN written whether book will-read not"
ii. *[$_{S'}$[$_{S}$[$_{NP}$[$_{AP}$[$_{NP}$ hemontor][$_{A}$[$_{A}$[$_{V}$ lekh] a][$_{COMP}$ bujhi]]][$_{N}$ boy]][$_{V}$[$_{V}$ poRben] na]]
"Hemonto-GEN written whether book will-read not"

The contrast between (46) and (48) points up the difference between a true satellite – a non-H daughter of H′ – and a non-satellite 'branch' like a non-H node J in a structure [$_{K}$ J H]. In (46), the AP modifying the head N is a satellite, and COMP Preposing fails to break into the NP. But in (48) COMP Preposing has broken into an AP + V sequence, legitimately, since the AP is not a true satellite in the structure [$_{S}$…AP V], there being no head node in this structure. Notice that this constitutes direct evidence supporting Hornstein's bar theory against Halitsky's (Hornstein 1975, Halitsky 1975). The derivation of (48) from (47) supports also the no-VP-node analysis of Bangla finite clauses adopted in this investigation.

(47) [$_{S'}$[$_{S}$[$_{AP}$[$_{AP}$ khub[$_{A}$ bhalo]][$_{V}$ khEle]][$_{COMP}$ ki]]
"very good plays whether"
'Does he/she play very well?'

(48) [$_{S'}$[$_{S}$[$_{AP}$[$_{AP}$[$_{AP}$ khub][$_{A}$ bhalo]][$_{COMP}$ ki]][$_{V}$ khEle]]]
"very good whether plays"
'*Does* he/she play very well?'

Although I have given a very task-specific formulation, (43), of the relevant condition on the application of movement rules, principle (43) could have been, in a more thorough investigation, shown to follow from Mark Baltin's (1978) 'landing site' theory of movement rules.

While on the subject of constraints on transformations, I wish to point out that the complementizer preposing rule also observes the Coordinate Structure Constraint first proposed by Ross (1967). Thus, (7) does not transform to (49).

(7) [$_{S'}$[$_{S}$ bap ma mara gEche][$_{CONJ}$ ba][$_{S}$ ghOr dor bheSe gEche][bujhi]]
"father mother dead have-gone or house door floating have-gone whether"
'Have their parents died or their property been flooded?'

(49) *[$_{S'}$[$_{S}$ bap ma mara gEchel][$_{CONJ}$ ba][$_{S}$ ghOr dor bujhi bheSe gEche]]
"father mother dead have-gone or house door whether floating have-gone"

Conjunctions, which figure in coordinate structures, deserve to be our next object of scrutiny, for some words seem to behave dually – to act as conjunctions in some places and as complementizers in others.

6.3 Conjunctions and Complementizers

The term Conjunction as used in current generative grammar abbreviates the traditional grammatical term Coordinating Conjunction. This abbreviation reflects the fact that Subordinating Conjunctions, traditionally considered the other subclass of Conjunctions, have ceased to exist as a unitary category. Some former Subordinating Conjunctions have been reclassified as Particles which take sentential complements. Others have been reclassified as Complementizers. With this reclassification, which has been completed under the auspices of generativism but was already under way in the work of the structuralist and later-traditional grammarians, concern for properties which the 'Coordinating Conjunctions' were thought to share with the 'Subordinating Conjunctions' has also been eclipsed by other concerns. I will now present evidence, that this eclipse has not been beneficial for the analysis of Bangla structure. Bangla shows obvious links between 'conjunctions' and 'complementizers'.

Three of the elements exemplified so far, the words *ba*, *je*, and *ki*, seem to need to be marked as capable of functioning both as conjunctions and as complementizers. While discussing the status of example (14) I pointed out that *kintu* has the meaning 'but' as a conjunction and 'however' as a complementizer. One must add to this list of dually categorized words *naki*, which means 'or rather...?' as a conjunction and 'apparently?' as a complementizer, and *to*, which means roughly 'in view of that' as a conjunction and 'obviously' as a complementizer.

In order to motivate the distinction between Complementizer and Conjunction as two concepts we must have some elements which are committed one way or the other. In my speech *bujhi* 'whether' is a committed COMP (in older dialects it has a CONJ use with the reading 'apparently') and *ki* 'whether' comes close to that status, restricted as its CONJ uses are to highly marked, and perhaps listed, environments (in older dialects its CONJ use, with the reading 'or', is quite productive). In all dialects, *Othoba* 'or', *kimba* 'or', *ebong* 'and', *karon* 'because', and various other elements are clearly committed Conjunctions. In view of the possibility that in some dialects there are no full-time complementizers, one may question the value of postulating a Complementizer category at all for a general grammar of the language. My response is that the postulation of COMP and CONJ as distinct classes makes it possible to state the otherwise unstatable fact that each clause is allowed at most one Conjunction plus at most one Complementizer. The need to make this statement would motivate the two categories even if every complementizer word had a homonym which was a conjunction, which, anyhow, does not hold of my dialect.

It is of course possible to translate all this into a system where +CONJ and +COMP are features. The dually classified words are then +CONJ +COMP and have a single lexical entry with, in some cases, branching to indicate, say, the meaning difference between *kintu* the CONJ and *kintu* the COMP. I find no difference between such a translation and my proposal of homonymy. But the translation makes for fewer lexical entries. So, I adopt it on account of Occam's razor.

One may raise another, more serious, objection against the theory of complementizers and conjunctions which I am suggesting. This objection is based on the fact that, for any word W which functions both as COMP and as CONJ, a given clause is allowed to contain at most one instance of W – a fact which I have not yet explained. I will return to this objection and slightly extend my model to cover the fact. At that point it will become crucial that dually functioning words each have a single lexical entry. But first I must articulate and defend the

present proposal in those respects in which it differs from Jayanti Chattopadhyay's (1976a) description of Bangla clause structure. She, too, studies *je*.

I assume the following procedure for the optional introduction of occurrences of the CONJ node into phrase-markers. First, the general schema (50) is applied. The output includes (51). A general convention for interpreting structures of the form (51) translates (51) into (52). The convention may be stated as (53).

(50) $X^m \rightarrow nX^m$, where *m* and *n* are integers, *m* representing the number of bars and *n* the number of copies of the X^m node in the output, and where X is a bundle of categorial features

(51) i. $[_{S'}$ 1S′]
ii. $[_{S'}$ 2S′]
iii. $[_{S'}$ 3S′]
iv. $[_{S}$ 1S]
v. $[_{S}$ 2S]

(52) i. $[_{S'}$ CONJ S′]
ii. $[_{S'}$ S′ CONJ S′]
iii. $[_{S'}$ S′ CONJ S′ CONJ S′]
iv. $[_{S}$ CONJ S]
v. $[_{S}$ S CONJ S]

(53) Interpret $[_{X^m}\ nX^m\ ]$ as $[_{X^m}\ X^m_1\ X^m_2 \ldots X^m_n]$ and then, into this structure, insert either nothing or *k* occurrences of CONJ (*k* =1 if *n* =1; otherwise *k* = *n* – 1) as follows. If *k* = 1, put $CONJ_1$ in $[_{X^m}\ \_\_\ X^m_1]$; if *k*>1, put $CONJ_i$ $(1 \leq i \leq k)$ in $[_{X^m} \ldots X^m_i \_\_\_ X^m_{i+1} \ldots]$

I continue to assume that the rule introducing COMP is (V-99-i), whose output undergoes (34), the optional rule of COMP Proposing. I claim that *je* 'that' enters both COMP and CONJ slots. I have to defend this claim against Chattopadhyay's claim that *je* is a COMP in her sentences (III-5-b) and (IV-1) repeated below as (54) and (55) respectively. For me, *je* is a COMP in (54) but a CONJ in (55).

(54) $[_S[_S$ kal je kOlej bOndo thakbe] amra agey e khObor peYechi]
"tomorrow that college closed will-be we already this information have-received"
'That the college will be closed tomorrow we already know'

(55) ram bujheche je kal kew aSbe na
"Ram has-understood that tomorrow anybody will-come not"
'Ram has understood that no one will come tomorrow'

The syntactic labels in (54) follow Chattopadhyay's analysis. Agreeing with her about *amra agey e khObor peYechi* not being a single constituent but not about *kal je kOlej bOndo thakbe* being an S rather than an S′, let us assign to (54) the structure (57), which uses PS rule (56), and let us assign to (55) the structure (58) which follows our analysis of conjunctions. I generate (58) directly from base rules and (57) via the deep structure (59), which undergoes COMP Preposing to yield (57).

(56) S → (S′) (PP) NP (PP) (NP) V

(57) [$_S$[$_{S'}$[$_S$[$_{PP}$[$_{PP}$ kal][$_{COMP}$ je]][$_{NP}$ kOlej][$_V$ bOndo thakbe]]][$_{NP}$ amra]
[$_{PP}$ agey][$_{NP}$ e khObor][$_V$ peYechi]]

(58) [$_S$[$_S$[$_{NP}$ ram][$_V$ bujheche]][$_{CONJ}$ je][$_S$[$_{PP}$ kal][$_{NP}$ kew][$_V$ aSbe na]]]

(59) [$_S$[$_{S'}$[$_S$[$_{PP}$ kal][$_{NP}$ kOlej][$_V$ bOndo thakbe]][$_{COMP}$ je]][$_{NP}$ amra]
[$_{PP}$ agey][$_{NP}$ e khObor][$_V$ peYechi]]

Chattopadhyay derives her structures (54)-(55) as in (60)-(61).

(60) i. [$_S$ amra agey [$_{NP}$[$_S$ kal kOlej bOndo thakbe][$_N$ khObor]] peYechi]
"we already tomorrow college closed will-be information have-received"

ii. result of applying (62), je-Complementizer Placement:
[$_S$ amra agey [$_{NP}$[$_S$ je kal kOlej bOndo thakbe][$_N$ khObor]] peYechi]
"we already that tomorrow college closed will-be information have-received"

iii. result of applying (63), S-Preposing:
[$_S$[$_S$ je kal kOlej bOndo thakbe] amra agey [$_{NP}$ e [$_N$ khObor]] peYechi]
"that tomorrow college closed will-be we already this information have-received"

iv. result of applying (64), Complementizer Postposing:
[$_S$[$_S$ kal je kOlej bOndo thakbe amra agey [$_{NP}$ e [$_N$ khObor]] peYechi] "tomorrow that college closed will-be we already this information have-received"

(61) i. [$_S$ ram [$_{NP}$[$_S$ kal kew aSbe na]] bujheche]
"Ram tomorrow anybody will-come not has-understood"

ii. result of applying (62), *je*-Complementizer Placement:
[$_S$ ram [$_{NP}$ [$_S$ je kal kew aSbe na]] bujheche]
"Ram that tomorrow anybody will-come not has-understood"

iii. result of applying (65), S-Postposing:
[$_S$ ram bujheche [$_S$ je kal kew aSbe na]]
"Ram has-understood that tomorrow anybody will-come not"

(62) *je*-Complementizer Placement (Chattopadhyay 1976a: 37), obligatory
S.D.: X – [S – (N)] – Y
S.C.: 1 2 3 4 ➔ *je*&2, 3, 4; & = daughter-adjunction

(63) S-Preposing (Chattopadhyay 1976a: 63), optional, follows (62)
S.D.: X – [S – (N)] – Y
S.C.: 1 2 3 4 ➔ 2@1, *e*, 3, 4; @ = sister-adjunction

(64) Complementizer Postposing (Chattopadhyay 1976a: 63), must follow (63) if (63) has applied; inapplicable otherwise

NP
S.D.: *je* – { ADV } – X
S.C.: 1 2 3 ➔ 0, 2@1, 3

(65) S-Postposing (Chattopadhyay 1976a: 67), global rule
S.D.: X – [S – (N)] – Y
S.C.: 1 2 3 4 ➔ 1, 0, 3, 4@2

The symbol + is used slightly diversely in Chattopadhyay's formalizations. I have taken the liberty of marking daughter-adjunction as & and sister-adjunction as @. In (63) I have also indicated that term 2 gets replaced by the determiner *e*, a fact which Chattopadhyay informally notes on page 29 but does not formalize. Note that this is the determiner /e/ and not the conventional "*e*" which marks an empty node.

At this stage I wish to show only that Chattopadhyay's analysis is wrong in assigning to the clause-initial *je* the status of a Complementizer instead of, as in my grammar, that of a Conjunction. Later I will return to other problems with her analyses.

My first argument rests on Imperative clause data.

(66) SOrkar nirdeS dEY ni je puliS bondider nirjaton koruk
"regime instruction gives not-PERF that police prisoners torture do-IMP"
'The regime had not given the instruction that the police should torture the prisoners'

(67) *puliS je bondider nirjaton koruk SOrkar e nirdeS dEY ni
"police that prisoners torture do-IMP regime this instruction gives not-PERF"

It is a fact about Bangla that a clause containing an Imperative form of a verb only tolerates the sort of occurrence of *je* that I would regard as a CONJ – i.e. an initial *je*. Chattopadhyay has no way to account for this fact. Her optional rule (63) followed by its compulsory sequel (64) derives (67) from (66) and thus predicts, incorrectly, that (67) should be a well-formed sentence. In contrast, my model can quite simply use a redundancy rule, such as the rule COMP ➔ [–[IMP…_]] – or something else based on the analysis of Imperative endings in my set-up for the verb inflection system – and thus prevent complementizers from being inserted into an S′ if the finite verb in it is in the imperative, predicting that (67) will be ill-formed. Chattopadhyay cannot correspondingly prevent her *je*-Complementizer Placement rule (62) from placing a COMP on an imperative-verbed clause, since she must generate (66) by doing just that. She has no way to generate (66) and not generate (67). Her analysis mistakenly equates the *je* of (66) with the *je* of (67).

My second argument more directly bears on the question of whether clause-initial *je* is a Conjunction. My analysis predicts that since each clause has only one CONJ slot, sentence (68), which tries to pack what in my terms are two conjunctions (*ebong* and *je*) into one slot, should be ill-formed, as it indeed is, in contrast to the well-formed (69), which my analysis allows since the occurrences of *je* in (69) count as COMP for my theory, being clause-non-

initial.

(68) *amra agey khObor peYechi je kal kOlej khulbe ebong je porsu khEla hObe
"we already information have-received that tomorrow college will-open and that day-after-tomorrow game will-be"

(69) kal je kOlej khulbe ebong porSu je khEla hObe amra agey e khObor peYechi
"tomorrow that college will-open and day-after-tomorrow that game will-be we already this news have-received"
'We already know that the college will open tomorrow and that the game will be played the day after tomorrow'

In contrast, Chattopadhyay must assign the same status to both (68) and (69). This prediction is incorrect. Therefore, her analysis is inadequate – again because of its failure to distinguish clause-initial *je* (for me, a conjunction) from *je* in other positions (for me, a complementizer).

Precisely what status her analysis will assign to (68) and (69) depends on her position with regard to the A-over-A principle, an issue she does not discuss. If she accepts A-over-A, her rule of COMP Placement will operate only on the entire conjoined structure *kal kOlej khulbe ebong porSu khEla hObe*, so that there will be exactly one *je* in the whole sentence. Thus, she will have derivations (70) and (71). (70) yields a surface structure which is the 'right way to say (68)', enabling her to predict that (68) should be ill-formed. But (71) yields a 'wrong way to say (69)', leading to the incorrect prediction that (69) too should be ill-formed. If Chattopadhyay rejects the A-over-A principle, she will get derivations (72) and (73). This time she will be right about (69) being well-formed but will incorrectly predict ungrammaticality for (68).

(70) i. amra agey [$_{NP}$[$_{S}$[$_{S}$ kal kOlej khulbe][$_{CONJ}$ ebong] [$_{S}$ porSu khEla hObe]][$_{N}$ khObor]] peYechi
ii. result of applying *je*-Complementizer Placement:
amra agey [$_{NP}$[$_{S}$ je[$_{S}$ kal kOlej khulbe][$_{CONJ}$ ebong][$_{S}$ porSu khEla hObe]][$_{N}$ khObor]] peYechi
iii. result of applying S-Postposing:
amra agey [$_{NP}$[$_{N}$ khObor]] peYechi [$_{S}$ je[$_{S}$ kal kOlej khulbe] [$_{CONJ}$ ebong][$_{S}$ porSu khEla hObe]]

(71) i. amra agey [$_{NP}$[$_{S}$[$_{S}$ kal kOlej khulbe][$_{CONJ}$ ebong][$_{S}$ porSu khEla hObe]][$_{N}$ khObor]] peYechi
ii. result of applying *je*-Complementizer Placement:
amra agey [$_{NP}$[$_{S}$ je[$_{S}$ kal kOlej khulbe][$_{CONJ}$ ebong][$_{S}$ porSu khEla hObe]][$_{N}$ khObor]] peYechi
iii. result of applying S-Preposing:
[$_{S}$ je[$_{S}$ kal kOlej khulbe][$_{CONJ}$ ebong][$_{S}$ porSu khEla hObe]]amra agey[$_{NP}$ e[$_{N}$ khObor]] peYechi
iv. result of applying COMP postposing:
*[$_{S}$[$_{S}$ kal je kOlej khulbe][$_{CONJ}$ ebong][$_{S}$ porSu khEla hObe]]
amra agey [$_{NP}$ e[$_{N}$ khObor]] peYechi

(72) i. amra agey[$_{NP}$[$_{S}$[$_{S}$ kal kOlej khulbe][$_{C0NJ}$ ebong][$_{S}$ porSu khEla hObe]][$_{N}$ khObor]] peYechi
ii. result of applying *je*-Complementizer Placement:
amra agey[$_{NP}$[$_{S}$[$_{S}$ je kal kOlej khulbe][$_{CONJ}$ ebong][$_{S}$ je porSu khEla hObe]][$_{N}$ khObor]] peYechi
iii. result of applying S-Postposing:
*amra agey [$_{NP}$[$_{N}$ khObor]] peYechi [$_{S}$[$_{S}$ je kal kOlej khulbe] [$_{CONJ}$ ebong][$_{S}$ je porSu khEla hObe]]

(73) i. amra agey[$_{NP}$[$_{S}$[$_{S}$ kal kOlej khulbe][$_{CONJ}$ ebong][$_{S}$ porSu khEla hObe]][$_{N}$ khObor]] peYechi
ii. result of applying *je*-Complementizer Placement:
amra agey [$_{NP}$ [$_{S}$ [$_{S}$ je kal kOlej khulbe][$_{CONJ}$ ebong][$_{S}$ je porSu khEla hObe]] [$_{N}$ khObor]] peYechi
iii. result of applying S-Preposing:
[$_{S}$[$_{S}$ je kal kOlej khulbe][$_{CONJ}$ ebong][$_{S}$ je porSu khEla hObe]] amra agey [$_{NP}$ e [$_{N}$ khObor]] peYechi
iv. result of applying COMP Postposing:
[$_{S}$[$_{S}$ kal je kOlej khulbe][$_{C0NJ}$ ebong][$_{S}$ porSu je khEla hObe]] amar agey [$_{NP}$ e [$_{N}$ khObor]] peYechi

I conclude that the data support my analysis, rather than Chattopadhyay's, of *je* in clause-initial position and in other positions. So, I now return to the more serious objection I mentioned – the objection based on the fact that (74), for example, is ill-formed, although my theory as it stands predicts that (74) should be fine.

(74) *[$_{S'}$[$_{CONJ}$ kintu][$_{S'}$[$_{S}$ tomar doy bhal lage][$_{COMP}$ kintu]]]
"but you-GEN yogurt good feels however"
'?But you like yogurt, however'

The questionable status of the English gloss in this case leads me to suspect that the unacceptability of (74) might have to do with semantic infelicity – that one may appeal to a principle like the Gricean conversational maxim about Quantity. But, just in case semantic inquiry shows my suspicion to be wishful thinking, let me offer a syntactic analysis which relies on the semantic properties of *kintu* in its two roles.

I posit an obligatory transformation (75) which, like all movement or deletion rules, obeys the principle of recoverability of deletion, and thus will insert a +CONJ item into position 2 only if the occupant of position 2 is identical to that of position 4 or if position 2 is unoccupied. Thus, (74) and (76) will both become (77) by this rule (note that the *e* shown in (76) is the conventional symbol *e* for 'empty').

(75) Conjunctionization (obligatory)
S.D.: X – CONJ – S – [+CONJ]
S.C.: 1 2 3 4 ➔ 1, 4, 3, 0

(76) *[$_{S'}$[$_{CONJ}$ *e*][$_{S'}$[$_{S}$ tomar doy bhal lage][$_{COMP}$ kintu]]]
"*e* you-Gen yogurt good feels but/however"

(77) [$_{S'}$[$_{CONJ}$ kintu][$_{S'}$[$_{S}$ tomar doy bhal lage]]]
"but you-Gen yogurt good feels"
'But you like yogurt'

The rule of conjunctionization will not make the mistake of moving one 'real conjunction' into the spot of another, because, thanks to the PS rule introducing conjunctions, no CONJ ever occurs at the end of an S′, and (75) will only become applicable in an S′ domain. Also, notice that, unlike the Infinitive Reanalysis rule that induced pruning of the VP node, which it deprived of its head, conjunctionization does not induce pruning of S′, since, if S′ has a head, its head is not COMP. (By bar notation conventions, S is the head of S′, though nothing of empirical use seems to follow from this fact.)

The application of (75) depends heavily on the fact that the lexical entries for certain complementizers are marked +CONJ as well as +COMP and the fact that the standard formal theory of lexical insertion, reinforced by the later research of Seegmiller (1974), requires insertion of the lexical entry together with its features into the phrase-marker. I am indebted to Eileen Fitzpatrick for drawing my attention to the consequences of this feature of the standard formalization of lexical insertion. To the extent that (75) is well-founded and supported by the data, its validity provides an argument for this feature of the standard formalization.

Here is an interesting piece of evidence for (75). The word *kintu* is not the only conjunction-complementizer which has one meaning as a conjunction and another meaning as a complementizer. The word *ba* is like that too. It means 'or' as a conjunction. As a complementizer, it always immediately follows an occurrence of [$_{EMP}$ i], and the *i-ba* sequence means 'indeed'. If, as we have decided, we are to have a single lexical entry for each conjunction-complementizer, then the entry for *ba* will have to have internal branching, with the conjunction subentry carrying the semantic interpretation 'or' and the subcategorization feature [–[EMP__]], and with the complementizer subentry carrying the reading 'indeed' and the subcategorization feature [+[EMP__]] (or, to be more specific, [+[*i* __]], if that degree of detail is called for). Now, when an entry comprises subentries, only one subentry at a time can be lexically inserted, since the subentries are 'distinct' from each other in the technical sense of distinctness. In such a set-up we can predict that *ba*, unlike *kintu*, will not undergo Conjunctionization, for its two subentries differ featurally. (In contrast, 'but' and 'however' are close enough to make the two *kintu*s non-distinct). We will not get a derivation of **ba kEno bolbe-i*? "or why? you-will-say-EMP" from *(ba) kEno bolbe-i ba?* "(or) why? you-will-say-EMP indeed" '(Or) why, indeed, will-you say it?', since an attempt to apply Conjunctionization will show that terms 2 and 4 of the S.D. match up with factors in the phrase-marker which are distinct – the recoverability principle will block the application of (75) here. There is some unclarity in the data at this point: *ba kEno bolbe-i ba?* is a borderline sentence (*kEno bolbe-i ba?* is fine). Perhaps the Gricean maxim of Quantity is relevant after all, even if it is not the only factor to play a role.

A somewhat subtle point of rule application procedure seems worth discussing in connection with the obligatoriness of (75). Since, like other present-day lexicalist linguists, I am assuming that movement rules are normally unordered, I should allow derivations in which the COMP (one which, say, happens to be +CONJ) preposes into the S before rule (75) is

brought to bear on the phrase-marker. Now, when (75) seeks to apply, its S.D. is no longer met, since the +CONJ complementizer now occurs inside rather than after the S. So the rule fails to apply, and we end up with an ill-formed structure like (78).

(78) *[$_{S'}$ [$_{CONJ}$ kintu][$_{S'}$ [$_{S}$ tomar[$_{COMP}$ kintu] doy bhal lage]]]
"but you-GEN however yogurt good feels"

One solution to this problem might be to claim that if the obligatory rule (75) fails to apply to a structure containing a CONJ preceding an S′ it assigns a star. In that case, though, many innocent strings like (79) will be starred.

(79) [$_{S'}$ [$_{CONJ}$ kintu][$_{S'}$ tomar ki doy bhal lage]]
"but you-Gen whether yogurt good feels"
'But do you like yogurt?'

A better solution that comes to mind immediately is to order obligatory rules like (75) before such rules as COMP Preposing, which are optional and can muddle the input to obligatory rules. But this is unsatisfactory if it remains a mere stipulation. Although I cannot now think of a way to derive this stipulation from some general principle of grammar, I do hope that such a general explanation can be given. For example, one might find that in general rules which are obligatory (and thus depart from the unmarked case – from optional application) always may, and sometimes must, be applied in strict linear sequence with respect to each other and to optional rules. I cannot advance such a principle, though, since I have no large corpus of 'secondary data' (well-founded rules that obey the principle) to support its postulation.

Another solution might be to invoke Grice's maxim of Quantity and to claim that (78) and the like are redundant. Indeed, (78) strikes me as bad style rather than bad grammar, like *ba kEno bolbe-i ba?* which clearly must not be ruled out by the grammar. In the absence of anything better, I accept this solution.

6.4 The Coordinating Couple /hOY…nOY…/

An opponent of my claim that some complementizers enter or move into conjunction positions may cite examples such as the following, urging that *hOY…nOY…* is a pair of conjunctions but co-occurs with what I must regard as the conjunction *je* and therefore proves that this *je* is not a conjunction.

(80) amra bollam je hOY ram jabe nOY SEm jabe
"we said that either Ram will-go or Shyam will-go"
'We said that either Ram would go or Shyam would go'

To this example offered in lieu of an argument I respond with a counterexample.

(81) porSu khEla hObe ebong hOY ram jabe nOY SEm jabe
"day-after-tomorrow game will-be and either Ram will-go or Shyam will-go"
'There will be a game day after tomorrow and either Ram will go or Shyam will go'

Since (81) has a bona fide conjunction *ebong* in the position where (80) has *je* 'that', I conclude that (80) deals no fatal blow to my view that *je* in clause-initial position is a conjunction.

This exchange of examples, of course, is no substitute for a genuine debate involving analyses which assign *hOY…nOY…* to a specific category. Unfortunately, no one has such an analysis. No other pair of words occurs in anything resembling *hOY…nOY…*'s range of contexts. In this respect Bangla differs from even such a closely related language as Hindi; Hindi has *aur…aur…* 'both...and...', *yaa…yaa…* 'either…or…', and *na to…aur na… ~ na to…aur na hi…* 'neither…nor…', which form a large enough system to warrant setting up a category for coordinating couples. One could point to Bangla *na…na…* as a coordinating couple in sentences like *na jani juddho, na jani Santisthapon* "not know war, not know peacemaking" 'I/we know neither war nor peace-making'. But *na…na…* turns up in a restricted range of environments and is associated with stylistic reordering (e.g. the Object-Verb reordering in the example given).

I will therefore not attempt an analysis of *hOY…nOY…*, but turn instead to elements that partly resemble it and do form a class – the class of Emphasizers. My reason for discussing them here is that they interact with complementizers.

6.5 The Emphasizers /i/ and /o/

(82) ram-o jabe SEm-o jabe
"Ram-EMP will go Shyam-EMP will-go"
'Both Ram and Shyam will go'

(83) ram-i jak ar SEm-i jak tomar tate ki?
"Ram-EMP go-IMP and Shyam-EMP go-IMP you-GEN that-LOC what?"
'Whether Ram goes or Shyam goes, what is it to you?'

The partial resemblance between the Emphasizers /i/ and /o/ and the coordinating couple *hOY…nOY…* becomes apparent in (82)-(83). So do three differences. First, the Emphasizers are enclitics, encliticized to the host element. Second, when an Emphasizer occurs paired, the same form is repeated: …i…i… or …o…o…. Third, the class of Emphasizers in the language has more than one member – it has two! – so that statements about Emphasizers are twice as significant as statements about 'coordinating couples'.

At most one emphasizer is allowed per word. Thanks to Joan Bachenko, who raised the question of how many emphasizers occur per phrase, we need not specially stipulate 'one emphasizer per word', since this can now be derived from the more general facts about emphasizers relative to phrases. These more general facts are as follows. If one defines 'phrase nodes' so as to exclude (among other things) S and S′, 'words' so as to exclude anything which cannot take a clitic, and 'satellites' so as to exclude everything which is a head, then one can state that every surface structure phrase which is either a daughter of an S or an NP daughter of an N^n (where *n* is some number of bars) is entitled to at most one Emphasizer and that this emphasizer occurs encliticized either to the phrase as a whole or to some word which must belong to or be the head of the phrase. These facts obviously entail that no word may have two or more emphasizers attached to it. Now we ask how to incorporate these facts into the

grammar of Bangla.

Once the emphasizer for a phrase has been introduced, at phrase-end and hence (Bangla being a left-branching language) in a position immediately following the head (if any – recall that S is an exocentric structure and has no head) of the phrase, it is a fairly straightforward matter to position the emphasizer correctly by applying local rule (84).

(84) EMP Preposing, optional
S.D.: X – X° – Y – EMP (X° being a zero-bar member of a bar system)
S.D.: 1 2 3 4 ➔ 1, 2+4, 3, 0

The application of rule (84) must obey the No New Satellite Condition (43) which ensures that the emphasizer does not jump over the head and the presumably universal condition on local rules which prevents local movement from extracting an item out of its mother phrase. The only intricacy worth paying attention to with respect to the conditions governing the application of (84) is, to my mind, the fact that the No New Satellite Condition, repeated below for convenience, does not prevent the derivation of (86) from (85), since the emphasizer's landing site is located not among the satellites or between the head and its satellites but inside the head itself.

(43) **The No New Satellite Condition**: Given [$_{H'}$ X_Y H Z_W], where H is a projection of a bar-notational feature bundle and is the head of H′ and where X, Y, Z, W (if non-null) are satellites of H, no rule may move material into either of the positions marked _.

(85) o tomake uttorTa [$_{V}$[$_{V}$[$_{V}$ bole][$_{V}$ debe]][$_{EMP}$ i]]
"(s)he you-OBJ answer-item tell-and will-give EMP"
'(S)he will definitely tell you the answer'

(86) o tomake uttorTa [$_{V}$[$_{V}$[$_{V}$ bole][$_{EMP}$ i]][$_{V}$ debe]]
"(s)he you-OBJ answer-item tell-and EMP will-give"
'(Why worry, since) (s)he will *tell* you the answer'

Turning now to the difficult problem of stating the rule which introduces the EMP into a phrase, we begin by noting that (43) will not govern the application of this rule, since the injunction of the No New Satellite Condition is against a new satellite being moved in from elsewhere in the phrase-marker, while a rule introducing EMP will bring it in *ex nihilo*. This, of course, tells us nothing about the rule itself. To begin with, we don't even know that the rule in question is an insertion transformation, or one rule rather than two.

Looking at the generalization that every surface structure phrase which is either a daughter of an S or an NP daughter of an NP is entitled to at most one emphasizer, we notice that S and NP, in view of our analysis of gerund phrases as NPs, are the two domains of lexical insertion where selectional relations exist between the 'predicate' (the verb in the sentence and the head noun in the noun phrase) and the 'arguments' (certain daughter phrases of the sentence or of the noun phrase). Along with this similarity between S and NP we must also note a difference. Since S, unlike NP, is the minimal and not the maximal projection of its bundle of categorial features (a bundle which we have been representing as +S), S has no head, and therefore cannot be said to have 'complements' or 'modifiers' in so far as these notions are

defined in relation to the notion 'head'. In the NP, on the other hand, since NP is the maximal projection of its categorial feature bundle (which Chomsky and Bresnan represent as +N –V, although other proposals exist), there is a head noun which takes complements and modifiers. In Bangla, since an NP is usually not allowed to contain a PP, the rule of thumb is that an AP in an NP is a modifier while an NP in an NP is a complement. (This claim, which could be defended elaborately, I ask you to accept here at face value. A detailed discussion of its merits would lead us too far afield and, at worst, be vitiated by basic unclarities stemming from the sorry state of the art of distinguishing complements from modifiers and specifiers in a language where the head is non-medial so that the distinction cannot, as in the existing proposals for English and French, be configurational.)

In terms of these points of similarity and difference between S and NP in relation to their daughters, we may state the generalization thus:

(87) Each non-modifier phrase which is the daughter of S or NP in surface structure may take at most one Emphasizer

It seems to me that one can most appropriately formalize (87) by postulating two transformations, an S-cyclic local rule (88) which places an emphasizer in adjunction to any daughter of an S (since S has no non-phrase daughters which are capable of taking a clitic, no disasters will happen) and an NP-cyclic transformation (89) which adjoins EMP to an NP descendant of an NP. You notice, no doubt, that (89) is wide of the mark, placing EMP as it does next to any descendant NP rather than only daughter NPs, and using as it does the notion of NP, and not that of 'non-modifier' which figured in generalization (87). Luckily for us, on account of coincidences in Bangla structure which it would take too long to explain, there are practically no empirical differences between (88)-(89) and (87), and the few places where differences do exist seem to bear out (88)-(89), although the crucial grammaticality judgments are insecure. But the problem remains partly open, in that (87) looks like a simpler hypothesis and should be formalizable as is, although current formalism in the generative tradition does not permit direct formalization of (87) since one cannot, in the S.D. of a rule, lawfully express the notions 'non-modifier' or 'daughter of NP'. That (88) manages to express the notion 'phrase daughter of S' is a partly fortuitous consequence of the system of categorial features assumed here (distinguishing S and S′ from the V-bar subsystem) and any version of the A-over-A condition. Presumably, someone will propose a modification of transformational theory to permit the notion 'phrase daughter of S' to be expressed without relying on such a coincidence and to permit the notion 'non-modifier daughter of NP' to be expressed. In this connection it is useful to recall Baker's (1973:254) remark that the fundamental difficulty in attempting to formalize Case agreement rules in generative terms is that the generative framework 'does not allow the traditional term *modifies* as one of the concepts that may appear in the statement of grammatical rules'. I can only leave the theoretical issue open, pending the construction of an adequate theory of a large range of grammatical relations (going beyond the narrow confines within which relational grammarians have been trying to carry out such an inquiry).

(88) Sentential EMP Placement (S-cyclic local rule)
S.D.: X – [–S] – Y
S.C.: 1 2 3 ➔ 1, 2+EMP, 3

(89) Nominal EMP Placement (NP-cyclic local rule)
S.D.: X – NP – Y
S.C.: 1 2 3 ➔1, 2+EMP, 3

In order to complete the examination of the distribution of EMP in large domains, let us observe some effects of the interaction of (88)-(89) with other parts of Bangla grammar. The single-quote glosses for (90)-(91) and (93)-(94) use the word *but* in an archaic sense, which I hope will not cause undue difficulty: I was hard put to think of a contemporary English expression that would do the same work.

(90) ram bhalo chobi-i aMke
"Ram good picture-EMP draws"
'Ram draws but good pictures'

(91) ram chobi-i bhalo aMke
"Ram picture-EMP good draws"
'Ram draws but pictures well'

(92) i. ram bhalo-i chobi aMke "Ram good-EMP picture draws"
ii. ram chobi bhalo-i aMke "Ram picture good-EMP draws"
'Ram draws rather well'

(93) ram bhalo chobi-i paThalo
"Ram good picture-EMP sent"
'Ram sent but good pictures'

(94) *ram chobi-i bhalo paThalo
"Ram picture-EMP good sent"
Intended reading: 'Ram sent but pictures well'

(95) i. *ram bhalo-i chobi paThalo "Ram good-EMP picture sent"
ii. *ram chobi bhalo-i paThalo "Ram picture good-EMP sent"
'Ram sent pictures rather well'

(96) ora oke eTa-o ante-i dilo na
"they him/her this-EMP bring-EMP let not"
'They just didn't let him/her bring even this'

(97) *eTa-o ante-i ora oke dilo na
"this-EMP bring-EMP they him/her let not"

(90)-(95) illustrate the difference between the AP *bhalo* 'good/well' as a constituent of S and as a constituent of NP. As S-daughter, *bhalo* is entitled to an Emphasizer. As NP-daughter, it is not. I have not indicated the constituent structure, for the occurrence of *good* in the English renderings amply indicates that *bhalo* is a modifier of a noun in (90) and (93) but that in the

sequentially similar (92-i) and (95-i) *bhalo*, since it bears an emphasizer, is forced into an 'adverbial' (S-daughter) function and thus renders (95-i) unacceptable because one does not speak of 'sending well' as one speaks of 'drawing (pictures) well'.

(96)-(97) illustrate the difference between the NP *eTa* in (96), where it is an S-daughter and is thus entitled to an emphasizer of its own, and in (97), where it is not the daughter of either an S or an NP and therefore is not so entitled. In (96), the deep structure VP *eTa ante* is broken up by the rule of Infinitive Reanalysis which adjoins the head V *ante* to *dilo na* and thus causes the VP node to be pruned, so that the NP *eTa* is entitled to undergo EMP Placement by (88) and the new V node *ante dilo na* can get an EMP by the same rule – an EMP which then gets adjoined to *ante* by EMP Preposing, rule (84). In (97), the deep structure VP *eTa ante*, having been preposed, remains intact, so that it, as an S-daughter, can take an EMP, which (84) can only adjoin to its head *ante*, but its internal NP *eTa*, being a VP-daughter and not an S-daughter, cannot take an EMP.

I return now to the main concern of the discussion of emphasizers in this chapter – their interaction with complementizers co-occurring in the same structure.

6.6 Emphasizers and Complementizers

The conjunction-complementizer *ba*, briefly discussed in section 6.3, is further exemplified below.

(98) ram kEno-i ba baRi jabe?
"Ram why?-EMP COMP home will-go"
'Why indeed should Ram go home?' ('should' or 'shall')

(99) ram baRi jabe-i ba kEno?
"Ram home will-go-EMP COMP why?"
'Why indeed *should* Ram go home?'

(100) ram-i ba kEno baRi jabe?
"Ram-EMP COMP why? home will-go"
'Why indeed should *Ram* go home?'

(101) ram baRi-i ba jabe kEno?
"Ram home-EMP COMP will-go why?"
'Why indeed should Ram go *home*?'

(102) ram-i kEno baRi jabe?
"Ram-EMP why? home will-go"
'Why should it be Ram who goes home?'

The contrast between (102) and (100) in form and meaning suggests the contribution that the complementizer *ba*, present in (100) but not in (102), makes to the meaning of the sentences where it occurs, but there seems to be no way to reduce this contribution to anything specifiable in English or in logical notation. The relative unacceptability of (103) and the like suggests that

at least in this case COMP Preposing is obligatory, or that there is a surface filter prohibiting the occurrence of *i ba* at the end of a clause – the filter being applicable whenever, through non-application of COMP Preposing or of some rule postposing an item past the *i ba* sequence, a clause-final *i ba* arises.

(103) ram kEno baRi jabe-i ba
"Ram why? home will-go-EMP COMP"
'Why indeed should Ram go home?' (mildly less well-formed than (98)-(102))

However, such a filter will not suffice to rule out the similarly marginal (104i-ii).

(104) i. ?ram kEno baRi-i ba jabe
"Ram why? home-EMP COMP will-go"
'Why indeed should Ram go *home*, of all places?'
ii. ?ram-i ba baRi kEno jabe
"Ram-EMP COMP home why? will-go"
'Why indeed should it be Ram who goes home?'

For the time being, then, I conclude that the principles determining the well-formed subset of the various permutations of (98)-(102) are an object of further study for which no obvious formal hypotheses come to mind within existing grammatical theory.

Other complementizers do not go so far as to require adjunction to a word with an emphasizer enclitic on it, but show a preference for occurrence next to a word with an emphasizer if there is a word of that sort in the sentence. (105-i) and (106-i) are well-formed, while (105-ii) and (106-ii) are marginal to varying degrees.

(105) i. ram tomake-o to kOthaTa bOle ni
"Ram you-OBJ-EMP COMP thing item tells not-PERF"
'Ram hadn't told YOU about it either'
ii. ??ram tomake-o kOthaTa to bOle ni
"Ram you-OBJ-EMP thing item COMP tells not-PERF"

(106) i. ram tomake-o ki kOthaTa bOle ni?
"Ram you-OBJ-EMP COMP thing-item tells not-PERF"
'Hadn't Ram told YOU about it either?'
ii. ?ram tomake-o kOthaTa ki bOle ni?
"Ram you-OBJ-EMP thing item COMP tells not-PERF"

However, the preference is not absolute:

(107) ram ki tomake-o kOthaTa bOle ni?
"Ram COMP you-OBJ-EMP thing-item tells not-PERF"
'Hadn't Ram told even YOU about it?'

(108) ram tomake-o kOthaTa bOle ni to
"Ram you-OBJ-EMP thing item tells not-PERF COMP"

'Ram hadn't told even YOU about it'

Of course, we can conclude nothing definite on the basis of (105)-(108), and the problem is not sufficiently central in this study to warrant going hunting for the exact answer. I conclude this general discussion with a rough guess as to the direction in which the answer may lie. As the formulation of Chattopadhyay's rule of COMP Postposing suggests, it is highly typical for a Bangla complementizer to occur just after the first major constituent of a sentence – although her analysis fails because such occurrence is typical but by no means the only option for complementizer positioning. And, as the formulation of my PS rule S′ → S COMP suggests, it is also typical for a Bangla complementizer to occur at the end of its sentence. Let us suppose that in some sense these two positions – sentence-second and sentence-final – are the 'typical', 'neutral' positions for complementizers in the language. Then any sentence in which the complementizer strays from these positions will be seen as using the untypical positioning of the complementizer for some purpose such as highlighting the constituent to which the complementizer is immediately right-adjoined. (105-ii) and (106-ii) are such sentences. But in (105-ii) and (106-ii) there is also an emphasizer element serving a similar purpose, that of highlighting the constituent *tomake*. Perhaps it is the attempt to highlight simultaneously the second and third constituents that leads to the ill-formedness of (105-ii) and (106-ii).

The remarks I have made so far have featured more problems than answers. There are further problems which one cannot begin to formulate within the current understanding of complementizers and emphasizers. Despite this profusion of problems, I hope that the examples clearly show that principles must exist which coordinate the positioning of emphasizers with that of complementizers – if only for the sake of the complementizer *ba* which only occurs after the emphasizer *i*, but perhaps also in relation to the less clear pattern of (105)-(108).

The future investigation of principles coordinating the positioning of emphasizers with that of co-occurring complementizers may become easier if we carry out at least a 'pilot study' based on an environment where the data pattern emerges with clarity. Compound verbs provide such an environment; we shall spare you the actual examples for reasons of space, providing only the schematic format of the admissible sequences. The generalization is that, although it is possible for either or both of the final clitic elements in a [$_V$[$_V$ compound verb] EMP COMP] sequence to prepose, if both of them prepose then they are always adjoined to the same verb and always in the order EMP COMP. As a result the sequences [$_V$V COMP V EMP V] and [$_V$V EMP V COMP V] never occur. And the sequence V COMP EMP is ill-formed regardless of surrounding structure. Acceptable sequences are, for a two-verb compound verb, V V EMP COMP, V EMP COMP V, V EMP V COMP (where EMP Preposing has applied but COMP Preposing has not), and V COMP V EMP (the other way round); for a three-verb compound verb, V V V EMP COMP, V EMP COMP V V, V V EMP COMP V, V EMP V V COMP, V V EMP V COMP, V COMP V V EMP, and V V COMP V EMP. How are we to formalize this state of affairs? To use the prevalent metaphor, it would be too 'costly' to somehow restrict the application of the syntactic rules (which, being syntactic rules, are optional) of COMP Preposing and EMP Preposing in such a way that they produce all and only the well-formed sequences at the level of surface structure. An appropriate filter is also difficult to construct if one wishes to handle the entire range of facts with one filter, although it seems useful to posit a filter *COMP EMP independently of the rest of the facts. I am inclined to place the descriptive burden on the semantic component. Interpretive rules, if we follow this course, accord special treatment to preposed EMP and COMP elements. They treat such an element as *focusing* on a

subconstituent. *Focusing* has the property that, operating on a given constituent, it must apply to no more than one of its subconstituents. Hence the uninterpretability of V COMP V EMP V and V EMP V COMP V. Since this move does not handle the ill-formedness of COMP EMP, I also postulate a filter *COMP EMP at the level of surface structure. The ill-formedness of V COMP V EMP V and V EMP V COMP V, then, results from the action of interpretive rules which apply to surface structure, and thus is not determined at the level of surface structure itself, but at the level of interpreted structure.

The above remarks can, as a 'pilot' study, help in future work on COMP-EMP co-occurrence only if one realizes that the *COMP EMP filter suffers from a serious limitation – that it misses a generalization. The problem, which I do not know how to overcome, is that some of the EMP COMP sequences are lexically enforced, in particular, those involving the complementizer *ba*. To drive this point home I will now mention occurrences of *ba* in environments other than 'immediately after *i* in a constituent question'. There are the expressions *hObe-o ba* "will-be-EMP COMP" and *ta hObe-o ba* "that will-be-EMP COMP" which both mean 'that may well be (the case)' and cannot be explained away as an idiom, since the meaning of the whole patently derives from the meaning of the parts, which are all synchronically versatile grammatical elements (as opposed to non-versatile elements like *cran* in *cranberry*): *ba* needs to be lexically specified as being able to occur in these expressions. And then there is the class of sentences with *ba* which are simplex finite clauses whose main verb is in the Simple Past (with the suffix /yl/). Examples follow.

(109) chelegulo ekTu aWaj korlo-i ba
"boy-PL a-bit noise made-EMP COMP"
'So what if the boys are making a bit of noise?'

(110) ajke bajare na gele-i ba
"today-OBJ market-LOC not went-EMP COMP"
'Why go to the market today?'

(111) ajke bajare na-i ba gele
"today-OBJ market-LOC not-EMP COMP went"
'Why go to the market today?'

(110)-(111) indicate that there is something else happening here, for the /na/ element occurs before its finite verb rather than after it. Sidestepping those complexities, we note again that *ba* must occur with the emphasizer. We now have three classes of examples where *ba* must immediately follow an emphasizer and where this sequencing must be stated in the lexical entry for *ba*. In these cases, then, if one were to construct an ungrammatical string like (112), it would be ill-formed on two counts. First, it would violate the lexical subcategorization; second, it would violate the *COMP EMP filter.

(112) *ajke bajare na-ba-i gele
"today-OBJ market-LOC not-COMP-EMP went"

I see it as a problem that the ill-formedness receives, redundantly, two explanations – in terms of the general grammatical fact formalized as a filter and in terms of the special lexical fact

formalized as a subcategorization feature. Since the filter and the feature both seem to be indispensable and well-motivated, we cannot remove the redundancy by jettisoning one or the other.

One might think that the problem could be solved by resorting to one of the means of expressing commonalities among statements in the grammar. Examples of such means are bar-notational schemata which, without replacing the specific phrase structure expansion for each categorial system, bring out the common features, and the information-theoretic type of generalization proposed by Jackendoff in a bid to account for the shared features of, say, *aggressive*, *aggressor* and *aggression* without postulating a fictive verb **aggress* which then especially fails to surface. But such means of expressing commonality are inapplicable to the case at hand, for they only work within a single component. The problem we are facing involves two components: the lexicon and the component of filters. It was Seegmiller (personal conversation, 1978) who pointed out to me this limitation of existing proposals for expressing generalizations based on irreducible grammatical statements.

Having pointed out the existence and intractability of this problem involving the *COMP EMP filter, I now return to the main line of inquiry.

6.7 Summing up about Complementizers

After providing 'primary' justification for the postulation of complementizers by showing that claims made on the basis of this postulation hold up under testing, this chapter has offered a series of detailed analyses involving the movement of complementizers and relating complementizers to two other categories: conjunctions and emphasizers. To the extent that these analyses are cogent, they provide 'secondary' justification for the postulation of a category of complementizers in the grammar of Bangla. Besides, the analyses of conjunctions and emphasizers offered in this chapter will come into play later in this investigation.

Returning to the concerns of chapter 5 which led to the need to justify and elaborate the complementizer proposal, recall that hypothesis (V-100), repeated below for reference, presupposed a definition of OPEN as 'able to occur in an S′ structure'.

(V-100) Determiners are lexically specified as ±OPEN. +OPEN Determiner entries are further specified as ±INTERROGATIVE or ±INT. +OPEN +INT Determiners begin with a /k/. +OPEN –INT or Relative Determiners begin with a /j /. For well-formedness, the value of the INT feature for all INT-specified Determiners in a clause must equal the value of INT for that clause's COMP.

As I noted at the time, that definition of OPEN was provisional and would need modification in the light of the extension of the notion of S′. In this chapter it has become clear that the rule expanding a COMP as ±INT must be optional, leaving room for complementizers which are unspecified for the INT feature. Thus, on one hand, there now exist instances of S′ in which the COMP is unspecified for INT, and, on the other hand, there was always the consideration that any S′ (even one with a COMP specified for INT) must be able to contain non-open determiners. It is therefore necessary to disengage the concept of OPEN from the properties of the S′ node, without losing the original idea that only in an S′ may an open determiner occur.

Accordingly, the feature OPEN now becomes an unanalyzed primitive as far as

syntactic analysis is concerned. On the one hand, a +OPEN determiner must be specified for the feature INT and a –OPEN determiner must not be specified as +INT. On the other hand, the feature specification +OPEN triggers interpretive rules to be proposed later. Thanks to the matching requirement stated in (V-100), an open determiner without a matching complementizer (or, a fortiori, without a complementizer at all) in the same structure leads to ill-formedness. Thus, open determiners will occur only in S′, without any special stipulation to that effect.

I have just suggested that the feature OPEN, while becoming opaque to syntactic analysis, must be taken to trigger interpretive processes and therefore will apparently receive some kind of definition or motivation in the semantic component. If the experience of recent decades is any guide, one risks unclarity or worse by leaving a construct defined only in semantic terms. In the case of the construct OPEN, I fortunately do not need to run this risk. It is possible to provide a morphological definition of a core set of Open words which also share the semantic property of openness I have in mind. I will begin chapter 7 with the relevant morphological analysis, for which the ground has been prepared by the study of emphasizers in this chapter. Once a formally, morphologically firm class of Open words is established, it becomes possible to argue for slight extension of the class so as to cover semantically and distributionally akin cases while retaining the morphological/semantic 'core' set as providing the primary justification for setting up a class. Such a move resembles the refinement Lyons (1969) has proposed of the traditional idea of linking the concept Noun to the class of words that name things or persons. Lyons feels that, while obviously the semantically based concept of nouns makes errors in predicting membership of the grammatically useful class of nouns, one may usefully (i) choose a set of words which are both distributionally nouns and semantically names and (ii) extend this core set to cover words which share the same distributional features. This procedure, to his mind, justifies retaining the 'notional' term Noun to designate the extended set. Likewise for the class of 'Open Words'.

CHAPTER 7

PHORIC WORDS: +OPEN AND –OPEN MEMBERS OF A CLOSED SET

7.0 Strategy and Tactics

7.0.1 Strategy

As promised at the end of the last chapter, I will now offer a morphological explication of the notion 'open word'. It will be seen that open words form a natural class which joins another such natural class – or a few other such classes, depending on details of analysis – to form, together, a larger natural class which I will call 'Phoric Words'.

The set of phoric words is a closed set in the sense that diachronically no new members are being added to it. As will become clear, there are morphological reasons for considering certain complementizers to be phoric words.

The analysis of phoric words offered in this chapter is purely formal and does not directly broach the way in which semantic properties of different sorts of phoric words interact to determine the special features of relative clauses, constituent questions, complement clauses, and yes-no questions. That larger syntactic and semantic inquiry will be left for the closing chapters of this volume.

7.0.2 Tactics

A table of phoric words will, perhaps, suffice to convince the reader that the category of phoric words is a morphologically sound one. A detailed grammatical analysis of some columns of this table will serve to make the point with more rigour.

Then, as usual, secondary justification for the analysis will be provided, by postulating a rule of Phoric Postposing in the domain of the NP.

Finally, the link with complementizers will be made, both by motivating the inclusion of some complementizers in the class of phoric words and by sketching the relation of certain non-complementizer phoric words to the Complementizer node in their S′.

7.1 The 'core' class of open words

A tabulation is the best way to introduce the notion and to demonstrate the need for it.

Open Words

K-Word	GLOSS	J-Word	GLOSS
kon	'which?'	je	'which'
ke	'who?'	je	'who'
kake	'whom?'	jake	'whom'
kar	'whose?'	jar	'whose'
kara	'who-PL?'	jara	'who-PL'
kader	'whom/whose-PL?'	jader	'whom/whose-PL'
kini (a facetious form)	'who-HON?' (facetious)	jini	'who-HON'
(the facetious form		jaMke	'whom-HON'
manifestly has a		jaMr	'whose-HON'
defective		jaMra	'who-HON-PL'
declension)		jaMder	'whom/whose-HON-PL'
ki	'what? (DET or NP)'	je	'what-DET'
		ja	'what-NP'
kake	'what-OBJ?'	jake	'what-OBJ'
kiSer	'what-GEN?'	jar	'what-GEN'
kiSe	'what-LOC?'	jate	'what-LOC'
kOkhon	'when?'	jOkhon	'when'
kOkhonkar	'of what time?'	jOkhonkar	'(the time) of which'
kObe	'on what day?'	jObe (archaic form)	
kObekar	'of what day?'		
konkhan(e)	'(at) what place?'	jekhan(e)	'where'
konkhankar	'of what place?'	jekhankar	'(the place) of which'
kotha(Y)	'where?'	jetha(Y) (archaic form)	
kothakar	'of what place?'	jethakar (archaic form)	
koto	'how much/many?'	jOto	'(as much/many) as'
kO (distributionally restricted)	'how many?'	jO	(highly restricted) '(as many as)'
kEno	'why?'		
kEmon	'of what sort?'	jEmon	'(the same sort) as'
kirO(ko)m	'of what sort?'	jerO(ko)m	'(the same sort) as'
konrOkom	'of which sort?'		

On the fence that divides belonging from not belonging to the table sit forms like *ki jonne* 'for what?', *ki karone* 'for what reason?', *je jonne* '(that) for which', *je karone* '(the reason) for which', which all appear to be two-word sequences semantically akin to the word *kEno* 'why?'. The items in the table all seem to be single words, although some of them have internal constituent structure; whether such structure is syntactic or not is a difficult theoretical question which I will beg by assuming that it is syntactic. It seems clear that the noun *rOkom*, optionally contracted into *rOm* only next to phoric words like *ki, je* etc., loses wordhood. So does the noun *khan*, which never occurs on its own, but only next to phoric words. Incidentally, one might propose for *khan* a morphological analysis which spells out *ki-khan-(e)* and *ki-khan-kar* as suppletively *kotha(Y)* and *kothakar* respectively. This proposal, in conjunction with the asymmetry of the pair *kon* and *ki* versus the solitary *je*, has the virtue of explaining the non-existence of *jetha(Y)* and *jethakar* in modern Bangla. And recall the discussion of the uncertain noun *ni* (as in *ji-ni*), one of the points that came up after the Zero Item View was adopted for the form *je Ø* in chapter 3. Note that the genitive suffix *-kar* , as in *kon-khan-kar* and *je-khan-kar*, only attaches to a noun fused with a preceding determiner (or, in other cases, obligatorily

devoid of any determiner – see chapter 10 for discussion) and unable to take an adjective, and even then only if lexically called for (thus, there is no **ki-rOkom-kar*). This consideration seems to suggest that *khon* of *kOkhon* and *jOkhon* should also be regarded as a noun of highly restricted distribution, which slightly increases the distributional range of *kO* and *jO*. I am not sure if *kOtokkhon* 'how long?' and *jOtokkhon* '(as long) as' count as single words (and thus belong in the table) or not, nor whether to regard this *kkhon* as a co-allomorph to the *khon* just mentioned.

Table (1) shows a few discrepancies and many common points between the K-Word column and the J-Word column. The overall moral one should draw is surely that there is a class of K-Words as well as a class of J-Words, both of which possibly belong to some larger class, and so forth. This much is simply the remaining modicum of love for taxonomy in linguistic research. Fortunately, we have more to go on in this case. There is an interesting morphological, distributional generalization which gives us the right to group all members of table (1) together into a natural class of what I call 'open' words. This generalization is that all these words, in contrast to non-open phoric words, are intolerant of the emphasizer /o/ 'even, also. Table (2) lists the phoric words which are obviously non-open. Table (8) will list the phoric words which I demonstrate to be non-open.

(2)

T-Word	O-Word	E-Word	Gloss for E-Word
Se(y)	o(y)	e(y)	'this'
Se	o	e	'this person'
take	oke	eke	'this person-OBJ'
tar	or	er	'this person's'
tara	era	era	'these people'
tader-	oder	eder	'these people-OBJ/GEN'
tini	uni	ini	'this person-HON'
taMke	oMke	eMke	'this person-HON-OBJ'
taMr	oMr	eMr	'this person-HON-GEN'
taMra	oMra	eMra	'these people-HON'
taMder	oMder	eMder	'these people-HON-OBJ/GEN'
ta	o	e	'this'
take	oke	eke	'this-OBJ'
tar	or	er	'this-GEN'
tate	ote	ete	'this-LOC'
tOkhon	tOkhon	Ekhon	'now'
tOkhonkar	tOkhonkar	Ekhonkar	'present'
Se(y)khane	o(y)khane	e(y)khane	'here'
Se(y)khankar	o(y)khankar	e(y)khankar	'of this place'
hetha(Y)	hotha(Y)	Setha(Y)	(archaic forms)
hethakar	hothakar	Sethakar	(archaic forms)
tOto	Oto	Eto	'this much/many'
tEmon	Omon	Emon	'of this sort'
Se(y)rO(ko)m	o(y)rO(ko)m	e(y)rO(ko)m	'of this sort'

Renderings for the T-Words and O-Words may be obtained by substituting 'the/the same/that' and 'yonder/that' respectively for 'this' in the glosses for the E-Words. Some of the items in (l) and (2) are familiar from the discussion of pronouns earlier in this inquiry. 'Phoric Words' refers to the combined table (1+2+8). 'Open Words' refers to table (1). The generalization exemplified in the following sentences is that non-open phoric words, but not open phoric words, may take the emphasizer /o/.

(3) era-o caY "these people too want"

(4) ora-o caY "they too want"

(5) jara pacche na tara-o caY
"who-PL are-getting not the-people too want"
'Those who are not getting it want it too'

(6) *jara-o pacche na
"who-PL-too are-getting not": ill-formed in any environment

(7) *kara-o pacche na
"who-PL?-too are-getting not": ill-formed in any environment

Since the overwhelming majority of Bangla words and phrases can take the emphasizer /o/, this is a fairly striking property of the class of open words and warrants setting them up as a class on this basis alone.

7.2 On K-Words and the Emphasizer /o/

I must now defend my claim that K-Words do not take the emphasizer /o/ against the counter-claim by lexicographer Rajsekhar Basu that they do. There is a class of words which one may notionally term 'indefinite words' and formally 'K-o-Words'. A tabulation of them is given in (8) below. This tabulation, compared to (l), makes it immediately obvious that an Indefinite Word is formed by adding an /o/ element to the corresponding Interrogative Word. It does not follow that this /o/ element is the emphasizer /o/. Jnanendra Mohan Das's 1916 dictionary, the most widely used large dictionary of the language, refrains from equating the Indefinite-forming /o/ with the Emphasizer /o/ – quite properly, to my mind. But Rajsekhar Basu's 1973 dictionary (actually first published in the fifties, but posthumously refurbished), the most widely used desk dictionary of Bangla, does equate the Indefinite-forming element /o/ with the Emphasizer. I need to defend my generalization against Basu's position.

(8)

K-o-Word	Gloss for K-o-Word	Corresponding K-Word
kono	'some/any'	kon
kew	'someone/anyone'	ke
kakeW, kawke	'someone/anyone-OBJ'	kake
karo, karur	'someone/anyone-GEN'	kar
kichu	'something/ anything'	ki
kichur	'some/any-thing-GEN'	kiSer
kichute	'some/any-thing-LOC'	kiSe
kOkhono	'some time/ever'	kOkhon
kobhu (archaic form)	'some time/ever'	kObe
konokhan(e)	'somewhere/anywhere'	konkhan(e)
kothaW	'somewhere/anywhere'	kotha(Y)
konorOkom	'of some/any sort'	konrOkom

While no one has, to my knowledge, argued in favour of Basu's position, several arguments supporting my position can be constructed.

As I mentioned in chapter 6, no word can take more than one emphasizer. As a result, the emphasizer /i/ does not occur immediately before the emphasizer /o/, nor vice versa. But the Indefinite-forming /o/ does occur immediately before the emphasizer /i/: *kono-i* 'any….at all', *kew-i* 'anybody at all', *kichu-i* 'anything whatsoever', etc. This indicates that the Indefinite-forming element is not the emphasizer /o/. This argument is not circular: a prima facie differentiation of the emphasizer /o/ and the indefinite-forming /o/ is possible because of

the semantic contrast.

While the emphasizers do exhibit some allomorphy – *i* and *y*, *o* and *W* – it is semi-automatic, which is why the transcriptions of emphasizer-bearing words given in this volume use *-i* and *-o* even when the phonetic form is *-y* and *-W*. The phones *o* and *W* never contrast in Bangla and the phone *i* and *y* very rarely do (it is only systemic considerations which motivate setting *o* and *W* up as the phonemes /o/ and /W/ in some classical phonemic analyses). In sharp contrast, the indefinite-forming morpheme /o/ can appear in its principal morphic shape *o* or *W* or in its non-principal morphic shape *u* or *w*, and the alternation between the high and mid shapes is morphologically conditioned. Compare *kew* and *kich-u*, from *ke* and *ki* respectively, with *de-o* (phonetically *deW* or *deo*) and *ghi-o* (phonetically *ghio*) from *de* (a proper name) and *ghi* 'clarified butter'.

Furthermore, the indefinite-forming /o/, like Case endings, induces the occurrence of special allomorphs of the stem it is suffixed to, as in *kich-u* from *ki* parallelling *kiS-e* from *ki*: but the emphasizers never induce occurrence of special allomorphs of any neighbouring stem.

When an emphasizer is encliticized to an inflected noun, it always follows the Case ending – it never precedes it: thus, from the proper name *de* we have *de-ke-o*, *de-r-i*, and the like, never *de-o-ke* or *de-i-r*. But the indefinite-forming element may either follow or precede the Case ending in the Objective and Genitive of *ke: ka-w-ke* or *ka-ke-W, kar-u-r* or *kar-o.*

Yet another factor distinguishing the indefinite-forming element from emphasizer behaviour is that, when an emphasizer attaches to a (C)VC word, the V of that word is realized as a long vowel (as it is expected to be if the word stands alone; cf. [pa:p] /pap/ 'sin' and [pa:pi] /pap-i/ 'sin indeed') rather than as short (as it would have been if the (C)VC element were a non-final stem morpheme preceding an inflectional or derivational suffix; cf. [papi] /papi/ 'sinner'). In contrast, when the indefinite-forming /o/ is attached to a (C)VC word, for example *kon* 'which', the vowel of that word is phonetically short rather than long: /kono/ is [kono] 'some/any'. Notice that the noun /kon/ [ko:n] 'angle' plus the emphasizer /o/ yields /kon-o/ [ko:no] 'angle, also'.

At the very beginning of the discussion of emphasizers in chapter 6 I gave examples of paired emphasizers, /i/…./i/... and /o/…./o/**...** Thus *ram-o jabe SEm-o jabe* means 'Both Ram and Shyam will go'. In contrast, the indefinite-forming /o/ does not occur paired; you cannot say the semantically quite plausible *kawke dekhi ni kichu dekhi ni* "anyone-OBJ (I/we)-see not-PERF anything-OBJ (I/we)-see not-PERF" 'I saw neither anyone nor anything' with the typical contrastive intonation of paired emphasizer expressions – one must say *kawke dekhi ni kichu dekhi ni* with the intonation of a paratactic sequence and the meaning 'I saw no one, I heard no one'.

Special plurals formed by reduplication, like *ke ke* 'which people?', *ki ki jiniS* 'what things?', *kon kon kOtha* 'which matters?', can take the indefinite-forming suffix to yield *kew kew* 'some people', *kichu kichu jiniS* 'some things', and *kono kono kOtha* 'some matters'. But forms like *ke ke, ki ki jiniS, kon kon kOtha*, when they take a bona fide emphasizer like /i/, do so only once – thus, *ki ki jiniS-i ba kinle?* 'What things indeed did you buy?', not *ki-i ki-i jiniS.*

Finally, the mechanisms of emphasizer placement and preposing make it impossible for a determiner element to take an emphasizer, but obviously the indefinite-forming /o/ is attached, among other things, to the determiner *kon* to yield *kono.*

This series of considerations makes it clear that the emphasizer /o/ and the indefinitizing /o/ must be regarded as two separate elements. I do not intend to deny that they both derive from the Sanskrit element /api/ which functioned both as an 'emphasizer' (if one dares to set up

such a category for Sanskrit without detailed analysis) and as an indefinitizing suffix. If we make the usual assumption that Sanskrit had just one /api/ with two functions, we have here a hitherto unsuspected case of split, which historical linguists will no doubt seek to relate to other diachronic processes in the Indic linguistic area.

Although the emphasizer /o/ and the indefinite-forming /o/ are demonstrably distinct morphemes, they do act as though they were the same morpheme for the purposes of the Isachenko-Aronoff haplology filter (Aronoff 1976) which prohibits a sequence of two identical affixes (for example, this filter prohibits *-sk-sk-* in Russian). Thus, although Bangla indefinite phrases in general may take the emphasizer /o/ – for example, *kono kono baMdor-o* 'some monkeys, too' – an indefinite phrase ending in the indefinitizing /o/ element itself cannot take the emphasizer /o/, so that, say, *kew kew-o*, which should mean 'some people, too', is ill-formed. This fact suggests that the haplology filter prohibits a sequence XY of affixes even if X and Y are distinct morphemes and only happen to have some allomorphs in common – a conclusion for which independent evidence comes from such well-known cases as English *his* (instead of *his's*, which is the form one would expect) corresponding to *hers*, which, however, differ from *kew kew-o* in that deletion of the second affix removes the offending sequence and renders the shorter form usable in the role of the longer form (i.e. *his* can do the work that *his's* would have done, whereas *kew kew* does not take over the work of the non-existent *kew kew-o*; this work simply remains undone). I have no explanation for this difference. I can only point out that the meanings that forms like *kew kew-o*, *kothaW-o*, *kOkhon-o-o* would have expressed, if they did exist, can be conveyed instead by using slightly longer paraphrases like *kono kono lok-o* 'some people, too', *kono jEYgate-o* 'even in any place', *kono SomOYe-o* 'even at any time'; whereas, in contrast, if *his* had not taken over the job of *his's*, the meaning of *his's* would have gone unexpressed.

To avoid giving the impression that I am building a theory on just one Bangla example, let me point out that analogous cases exist for the element /i/ which occurs in words where it cannot be regarded as a syntactically introduced emphasizer; for example, in the forms *ey, oy, Sey*, which are variants of *e, o, Se* 'this, yonder, that' in table (2), or in the form *praYi* 'often', which can no longer be synchronically derived from *praY* 'almost', although there was no doubt a time when the meanings of *praY* and *praYi* were compositionally related. Well, there is no semantic reason why *praYi* 'often' should not take an emphasizer /i/ and yield *praYi-i*, which would have meant 'often indeed'; but *praYi-i* is in fact ill-formed, and, again, nothing takes the place of *praYi-i* the way *his* takes the place of *his's*.

Thus, although some unclarities remain, it seems reasonable to conclude that the ungrammaticality of *kew kew-o* and the like does not constitute data for an argument supporting Rajsekhar Basu's view (that the indefinite-forming /o/ is identical to the emphasizer /o/) against my view (that these elements are distinct from each other). And I have provided many arguments supporting my view against Basu's. Thus, I take it as established that indefinite forms are not an exception to my generalization that K-Words and J-Words do not take the emphasizer /o/. Thus, the notion of a natural class of open words including all and only K-Words and J-Words, a notion based on this generalization, emerges intact from the foregoing discussion of indefinite forms. Notice that, in view of the well-formedness of *kono kono baMdor-o* 'even some monkeys' and the like, indefinite phrases must be regarded as capable of taking the emphasizer /o/ in principle, suggesting that the inability of indefinite words to take the emphasizer /o/ must be attributed to the haplology filter: therefore, indefinite words must not be included in the class of open words. The crucial contrast is between *kon kon baMdor-o,*

which would have meant ‘even which monkeys?’ and is ill-formed, and *kono kono baMdor-o*, which bears the interpretation ‘even some monkeys’ and is well-formed.

7.3 Phoric Words as a Natural Class

So far, I have merely asserted that the class of phoric words, i.e. the contents of the open word table (l) and the non-open word tables (2) and (8), is morphologically a natural class. I have appealed to the reader’s sense of morphological patterning in making this assertion. I will now go further and say that the (admittedly not fully systematic) correspondences among the six columns in tables (l), (2), and (8) are found only among these six columns and not between any of these columns and any other forms in the language. To put it differently, the class of Phoric Words shown in these three tables is a closed set within which paradigmatic correspondences exist.

So far, so good. Now we must justify setting up a class of phoric words. This justification must have nothing to do with the morphological criteria which have been used in setting up the class in the first place. Focusing on the determiners in the phoric tables we find that these determiners occur at various points in the structure of an NP. In particular, they occur in all the positions marked __ in [$_{NP}$ (NP) __AP__ AP__N]. Since a given NP must not contain more than one determiner, the fact that this determiner may occur in one of these positions seems to need to be expressed by means of a movement rule.

(9) apnader je dirghomeYadi bipul khoti
“your-HON-PL which long-term immense loss”
‘your long-term immense loss which’

(10) apnader dirghomeYadi je bipul khoti
“your-HON-PL long-term which immense loss”
‘your long-term immense loss which’

(11) apnader dirghomeYadi bipul je khoti
“your-HON-PL long-term immense which loss”
‘your long-term immense loss which’

If one assumes, as seems natural, that (9) is the untransformed deep structure, then, assuming the internal structure NP DET AP AP N for the NP (9), we can state the movement rule initially as (12).

(12) S.D.: DET – X – Y
S.C.: 1 2 3 ➔ 0, 2, 1+3

Given the usual convention of adjunction, (12) will apply only when Y is non-null and can therefore take a left adjunct. Thus, (9) will have an alternative surface structure where *je*, instead of being a sister of *dirghomeYadi* and *bipul*, is adjoined to *dirghomeYadi*. In (10) and (11), *je* is adjoined respectively to *bipul* and to *khoti*.

Now, (12) as stated is too strong, in that not everything which functions as a

'determiner' seems to undergo this sort of postposing. A clear case, in my speech, is the determiner *prottek* 'each'; it occurs in structures like (13), where it is initial, but not in structures like (14) or (15), where it is medial.

(13) prottek dirghomeYadi biraT porikOlpona "each long-term large project"

(14) *dirghomeYadi prottek biraT porikOlpona "long-term each large project"

(15) *dirghomeYadi biraT prottek porikOlpona "long-term large each project"

Thus, although some determiners which are not phoric elements do undergo (12), in view of the data about *prottek* it seems necessary to work temporarily with a relatively limited rule (16) which is exceptionless, as transformations should be; we leave for later work the task of determining the conditions under which non-phoric determiners can sometimes undergo an analogous process.

(16) Phoric Determiner Postposing
S.D.: [+DET +PHOR] – X – Y
S.C.: 1 2 3 ➔ 0, 2, 1+3

I don't want to make this conservative move too hastily. It is necessary to point out that another option, which I have no strong reason to rule out, would be to decide instead that *prottek* is not a determiner. Indeed, *prottek* does also occur in positions where Numeral elements occur, for example in expressions like *prottek Ta boy* "each item book" 'each book', with the Denominator *Ta* – cf. *paMc Ta boy* "five item book" 'five books'. And the possibility of using *prottek* without a Denominator, as in (13), may reflect the same forces which are at work in expressions such as *Ek dirghomeYadi biraT porikOlpona* "a long-term large project", which deploys the Numeral *Ek* rather than the Numeral-Denominator sequence *Ek Ta*, with a slight difference of meaning ('a' rather than 'one'). However, the problem with this option is that the result of substituting *Ek* for *prottek* in (14) and (15) is well-formed: *dirghomeYadi Ek biraT porikOlpona* "long-term a large project", *dirghomeYadi biraT Ek porikOlpona* "long-term large a project'; why the status of these forms differs from that of (14) and (15) would remain unexplained if *prottek* and *Ek* were thought to belong to the same category NUML. We see, then, that the option of regarding *prottek* as a NUML rather than (instead of 'in addition to') a DET is not without its problems. Of course, it is always possible that *prottek* has other special features.

I should also point out that the data are less clear-cut than one might wish. (14) is only about half as unnatural as (15), though it is significantly worse than (13). I don't know why this difference exists.

Bearing these problems in mind, I observe that (16), while it may not be general enough, is certainly the most uncontroversial generalization possible for facts like (9)-(11). This seems to me to provide some evidence for the natural class of Phoric Words – for the feature [Phoric].

Notice that the rule of Phoric DET Postposing does not violate the No New Satellite Condition, since the DET is already inside its nominal domain and thus does not count as breaking in. Notice also that the rule requires that, in an NP that immediately dominates NP

(DET) AP N, the (DET) AP N sequence forms a constituent, which one might call N′ or NOM; for, if the domain of Phoric DET Postposing had been NP rather than N′, the rule would fail to apply to a DET in the environment NP __ AP N. The domain of (12), then, is N′.

In principle, I would expect to see similar things happening in the case of Phoric Determiners of Adjective Phrases. I have not been able to find or construct any plausible examples where an AP has the structure __ AP __ AP__ A, where we might test whether DET may occur in all the sites marked __. So at least there are no counterexamples to (16) in its maximally simple current formulation, which applies to DET in AP as well as to DET in N′. The statements in the preceding paragraph, then, should be modified to the extent that, throughout, N should read +N; A and N are +N categories and V and P are –N categories.

7.4 The Phoric Complementizers

The two complementizers which are most crucial to the argument of this work — *je* and *ki*, which one may gloss as 'that' and 'whether/or' respectively — have morphic shapes which warrant their inclusion in the natural class of phoric words. So does the complementizer *to*, which roughly means 'as is known'. They belong, respectively, in the J-Word column, the K-Word column, and the T- and O-Word columns of tables (1) and (2). The Complementizers *je* and *ki* have exactly the same shapes as the words *je* and *ki* already in those tables, and the Complementizer *to* has the shape, almost, of the *tO-* of *tO-khon* 'then' (which is a T-Word that also doubles, syncretistically, as an O-Word) and *tO-to* 'that much/ that many' (a straightforward T-Word). Turning from form to meaning, *ki* as a complementizer, like *ki* as a phoric determiner, is an interrogative element, and *to*, like the bona fide T-Words and the bona fide O-Words, can respectively refer to an antecedently known state of affairs and point to an easily perceptible state of affairs. A demonstration that the complementizer *je* is semantically similar to the bona fide J-Words will have to wait until we can give a detailed account of the relevant semantic processes – but we may observe right away that both the bona-fide J-Words like *jekhane* and the complementizer *je* somehow relate the clause in which they occur to the following clause in sentences like (17) and (18) below.

(17) o *je* jabe ami ta jantam
"(s)he that will-go I that knew"
'I knew that she/he would go'

(18) ram *jekhane* jabe SEm Sekhane age gEche
"Ram where will-go Shyam there before has-gone"
'Where Ram will go, Shyam has been before'

Thus, on grounds similar to those which initially motivated the setting up of the class of phoric words, one can motivate the inclusion of the complementizers *je*, *to*, and *ki* in this class. And, again, one must seek secondary justification for the decision to take this semantic-morphological correspondence seriously – the decision to regard the complementizers *je*, *to*, and *ki* as phoric. Such justification will be offered in chapter 9.

7.5 The COMP node and Phoric non-Complementizers

All the open non-complementizer Phoric words listed in (1) have a special relation in my theory to the COMP node in their S′. However, as examples given so far show, most (perhaps all) clauses containing open phoric words lack a visible surface complementizer. I have not yet justified the postulation of a COMP node (and thus of an S′ rather than S structure) for relative clauses and constituent questions. All I have done is show that the notion of a COMP node is a viable one and that two tangible occupants of the COMP slot, *je* and *ki*, themselves belong with the open words in table (1).

I now need to develop the part of my theory which deals with COMP in clauses containing non-COMP open words – relative words or interrogative words – from (1). The following discussion assumes the result, obtained in chapter 3, that those items in (1) which are not wholly DET nevertheless contain a relative/interrogative DET plus some kind of fragmentary or zero noun (the 'zero noun' case of this assertion was elaborately discussed in chapter 3; only for the fragmentary noun /ni/ in honorific forms was the fragmentary noun case touched upon; table (1) contains several further examples of fragmentary nouns, as noted immediately after (1)). However, this assumption is not quite crucial at the present stage. One might just as well assume that the +INT or –INT feature is associated with the entire word in cases like *kon-khan-e* 'where?' or *je-khan-e* 'where'.

I will compare three hypotheses. The No Complementizer Hypothesis (NCH) asserts that relative clauses and constituent questions are not S′ but S. The Complementizer Transfer Hypothesis (CTH) asserts that in deep structure the +INT Complementizer of a constituent question is *ki* while the –INT Complementizer of a relative clause is *je*; that in deep structure the non-complementizer open word slots are left unspecified for the feature INT; and that a COMP Transfer rule copies the K/J morpheme (in effect the INT specification, + or –) on to every non-COMP open word slot in the clause and vacates the COMP. The Semantic Matching Hypothesis (SMH) involves a semantic process construing the open words in a clause with the COMP of that clause in terms of the feature INT.

The least tenable position, NCH, differs from SMH and CTH in that it cannot prevent a clause from containing both a relative word and an interrogative word. But such clauses – e.g. *kon khap theke je tOloar berolo* "which? sheath from which sword emerged" '(the sword) which emerged from which? sheath' – are ungrammatical. SMH predicts this, since its semantic rule will find at least one of the open non-COMP words incompatible with the COMP of the clause which will be specified either as +INT or as –INT, but not both. CTH makes the same prediction too, since its CT rule will copy either the K or the J morpheme (but not both) on to non-COMP open word slots in the clause, so that a clause containing both sorts of open word will not be generated by the syntax. We must, therefore, reject NCH.

Over a large range of data, CTH and SMH have identical empirical consequences. I have found two sorts of crucial examples which choose between them. Both choose SMH. The first crucial phenomenon is the lexical specificities of the J-Word column and the K-Word column. There are no relative words corresponding to *kEno* 'why?' or *kObe* 'on what day?/ in what year?...'. But CTH requires perfect congruence between the relative system and the interrogative system. Therefore SMH, which drops this requirement, is better. The second set of crucial data has to do with a special type of clause which contains, occasionally, J-Words or K-Words. (19) gives examples of such clauses. (20), for the sake of contrast, exemplifies open-word-containing clauses which are more 'normal' in that they do not tolerate overt

complementizers, while the clauses in (19) do. The data in (19) are compatible with SMH but not with CTH; only CTH makes the false prediction that a clause containing a non-complementizer open word is bound not to contain an overt complementizer word.

(19) a.
M M MR L L L
SEm kEno-i *ba* phirbe? (pitch markings: L low, M medium, H high, R rising, F falling)
"Shyam why?-EMP COMP will-return?"
'Why, indeed, would Shyam return?'

b.
M H H M HF
kEno *je* phirle!
"why? COMP returned"
H L L L MMR
'Oh, *why* did you return?'

c.
H M H H H M HF
ki muSkile-i *na* poRtam!
"what? problem-LOC-EMP COMP would fall"
'What a problem it would (have) be(en) for us!'

d.
ja-o *ba* baki chilo Ekhon tar ar kichu-i ney
"what-EMP COMP left was now that-GEN any-longer anything-EMP remains-not"
'Of the little that remained, there is nothing left any more'

e.
jOto-i ceSTa koruk <u>na</u> (kEno), kaj SeS hObe na
"however-much effort do-IMP COMP (why?), work finished will-be not"
'However hard they may try, the work won't be finished'

(20) f.
jara pare na tara SekhaY
"who-PL can not they teach"
'Those who can't, teach'

g.
*jara to pare na tara SekhaY
"who-PL COMP can not they teach"

h.
*jara kintu pare na tara SekhaY
"who-PL COMP can not they teach"

i.
H H M MR
kEno phirle?
"why? returned"
'Why did you return?'

j.
*kEno to phirle?

"why? COMP returned"

k. *kEno kintu phirle?
"why? COMP returned"

CTH would fail to account for the data in (19), since it would begin by associating clause (19-a)-(19-c) and the relative clauses in (19-d) and (19-e) with deep structures featuring the COMP *ki* in the case of (19-a-c) and *je* in the case of (19-d-e) and then performing COMP Transfer to complete the specification of the items *kEno, ki, ja,* and *jOto*; after these operations, the thus vacated COMP node would have no way of acquiring new occupants like *ba, je,* or *na* on the way from the output of COMP Transfer to the end of the transformational derivation, since lexical insertion would have stopped taking place at an earlier point in the derivation. Notice that the surface *je* of (19-b) cannot be taken as a remnant[9] of the postulated deep structure *je* of relative clauses, since (19-b) is a constituent question and not a relative clause. On the other hand, if a proponent of CTH sought to avoid this difficulty by inserting the appropriate complementizers *ba, je, na* at the level of deep structure in (19), then, given that a COMP slot can contain only one complementizer at a time, there would be no way to derive *kEno, ki, ja,* and *jOto* in (19) by COMP Transfer, and thus no way to derive the full surface structures for (19).

SMH can easily account for (19). A proponent of SMH need posit nothing beyond the following lexical insertion options: *ba* can enter a +INT COMP slot following an EMP *i*; *ba* can enter a –INT COMP slot following an EMP *o* following a relative word (this statement is a slight oversimplification, in ways which are irrelevant here); the *je* which occurs in clauses like (19-b) can enter a +INT COMP slot; the *na* of (19-c) can enter a +INT COMP slot following an EMP *i*; the *na* of (19-e) can enter a –INT COMP slot following an Imperative verb with at least one occurrence of the EMP *i* prior to the verb in the clause.

Much the same would need to be stated in any grammar of Bangla. SMH, which posits no transformational relation between the complementizer and the relative/interrogative words elsewhere in the clause, needs nothing more than the above lexical insertion statements to handle the data in (20). The interpretive process of semantic matching associated with the Semantic Matching Hypothesis works just as well in (20-f) and (20-i), where the ±INT COMP is empty, as in (19), where the ±INT COMP has a visible occupant.

The problem which (20) poses for SMH is that, unlike CTH, SMH does not predict its ungrammaticalities by a general principle to the effect that relative clauses and constituent questions contain no overt complementizer word. Having no such general principle, SMH has to depend on specific lexical markings which prevent complementizer words like *to* and *kintu* from entering a ±INT COMP slot. This is obviously somewhat unsatisfactory. One way out of this difficulty is as follows.

[9] [A 2020 note:] This must be one of the few pre-Kaynean occurrences of the term 'remnant' in the generative literature. Carrying no theoretical weight, but worth flagging for bibliographic reasons.

7.6 Exclamations as a sentence class

Intonational and distributional motivation exists for setting up a sentence class called 'exclamations'. Comparing the intonation patterns in (19) with that of (20-i), one notices that the 'genuine' question (20-i) has a rising terminal contour, while the 'rhetorical questions' and/or 'exclamations' in (19) have falling terminal contours indicated in my notation either as F for Falling or in terms of a sequence going from non-Low to Low as in (19-a), M M MR L L L. Furthermore, (19-e) illustrates a distributional property which distinguishes exclamatory from other relative clauses: the relative clause in (19-e) contains a verb in the Imperative, unlike any non-exclamatory relative clause. Exclamations include sentences containing neither relative nor interrogative words, like (21).

M H HF
(21) jacchi to
"am-going COMP"
'I *am* going!' (annoyed at someone rushing me)

Other sorts of distributional motivation for a class of exclamations become apparent in (22).

(22) a. ja-o ba dirghomeYadi kaje hat dite partam ebar ar ta-o parbo na
"what-EMP COMP long-term task-LOC hand to-give could this-time any-longer that-EMP will-be-able not"
'Now I won't be able to undertake even the (few) long-term tasks I might have undertaken'

H L H M H H H HF
b. ki darun chobi tuleche!
"what? tremendous photos has-taken"
'What tremendous photos (s)he has taken!'

H MM ML LF
c. koy pherot diyeche!
"where? back has-given"
'(S)he obviously hasn't returned it'

M M M H L L L
d. aditto koy phireche!
"Aditto where? has-returned"
'Obviously Aditto hasn't come back'

In sentence (22-a), *ja* is used as a DET in a DET AP N noun phrase; in (22-b), *ki* is used as a DET in a DET A adjective phrase; in (22-c) and (22-d), *koy* is used in the sense of 'where?' with exasperation. In non-exclamatory sentences, *ja* and *ki* are not so used, and *koy* is only used in present tense clauses with no overt copula (e.g. *aditto koy?* 'Where (is) Aditto?'); the word for 'where?' in non-exclamatory environments is, as in (l), *kothaY* or *konkhane*.

Suppose we set up a class of Exclamations as a sentence type along the lines indicated

above. This class will subsume sentences which contain J-Words, sentences which contain K-Words, and sentences which contain neither. I shall call J-Word-containing and K-Word-containing Exclamations respectively Exclamatory Relative Clauses and Exclamatory Constituent Questions. I shall call non-Exclamations containing J-Words and K-Words respectively Real Relative Clauses and Real Constituent Questions in the following discussion. Later on, I will drop the word 'real'.

We are now in a position to argue that Real Relative Clauses and Real Constituent Questions never do contain an overt COMP, and that therefore the kernel of truth in CTH can be saved even if we choose to adopt instead the Semantic Matching Hypothesis in order to handle the facts about Exclamatory Relative Clauses and Exclamatory Constituent Questions. All the questions in (19) have the intonational characteristics of Exclamatory Constituent Questions rather than of Real ones; and, the relative clauses in (19) also behave like Exclamatory Relative Clauses, 19-e) especially so since it has an Imperative verb unlike Real Relative Clauses. Therefore, let us postulate a general principle ruling out surface complementizers for Real Relative Clauses and Real Constituent Questions while allowing some lexically marked complementizers to appear in Exclamatory Relative Clauses and Exclamatory Constituent Questions. Since the existence of the Exclamatory clauses makes it impossible to sustain the Complementizer Transfer Hypothesis (the strongest possible formulation of our 'general principle') as a comprehensive theory of relative clauses and constituent questions, our new general principle will need to be couched in semantic terms in relation to the Semantic Matching Hypothesis. The exact formulation of this principle will depend on an adequate semantic description of exclamatory constructions. At present, not even the basic principles of the semantics of exclamation are known. So, we will have to make do with an imprecisely delimited idea of what 'Real' Relative Clauses and Constituent Questions are at the semantic level and state that such 'Real' clauses are semantically processed in a way which requires that there be no actual occupant in the COMP slot (+INT or –INT) of such a clause.

The need to relegate the prohibition against overt complementizer words in real relative clauses and real constituent questions to the semantic level means that the syntax will generate the ungrammatical sentences in (20), for example. We must eschew lexical restrictions preventing such sentences, since we have agreed to state the restriction only once, as a generalization, at the semantic level. Therefore, our semantic description of clauses containing J-Words and K-Words now has the task of accounting for the ungrammaticality of such sentences by assigning them a derivation which is ill-formed for non-syntactic reasons.

Notice also that it is the semantics and not the syntax which must account for the ill-formedness of clauses like *kon khap theke je tOloar berolo* which would mean (if well-formed) '(the sword) which emerged from which? sheath' – clauses which mix relative with interrogative words. Since we have rejected the Complementizer Transfer Hypothesis which would rule such clauses out syntactically, we must fall back on the semantics.

To sum up, we need an intricate semantic theory in order to handle the facts about 'real' relative clauses and constituent questions, since the existence of the exclamatory relative clauses and constituent questions (which we do not know how to handle semantically) forces us to move away from a transformational hypothesis along the lines of the Complementizer Transfer idea.

Chapters 8 and 9 construct a semantic theory of the sort required.

Chapter 8

RELATIVE, INDEFINITE, AND CONSTITUENT-INTERROGATIVE

8.0 Strategy and tactics

This chapter analyzes the grammar of Bangla relative clauses, indefinite phrases, and constituent questions. Some aspects of this analysis presuppose familiarity with Particle Phrase structure. Accordingly, this chapter begins with an analysis of the latter. I first describe those PPs where P governs NP or AP and then move on to those where P governs S. Some of the latter structures which I examine have relative and interrogative words inside the S. I discuss the terminological question (of possibly substantive significance) as to whether such PPs should be called 'complement clauses', given that they often translate what grammars of other languages do regard as 'complement clauses'; I give my reasons for not using the term 'complement clauses' for these PPs.

I consider next the problem of the syntactic relation between relative clauses and their heads and conclude that relative clauses in some conditions are and in other conditions are not part of a constituent which includes just the relative clause and its head.

Then I discuss the semantics of relative clauses and propose a theory which links each relative phrase to an antecedent phrase, noting that this linkage can be formalized either in terms of a (1, k)-quantifier and a unique-existential quantifier or in terms of a grammatical notion of bound anaphora.

In a bid to support the latter alternative against the former, I propose a semantic description for indefinite expressions which makes use of the idea that semantic interpretation at least begins by going only halfway to meet logic. This description of indefinites directly entails a description of interrogatives, which I then elaborate on the basis of the Semantic Matching Hypothesis of chapter 7.

This elaboration, when extended to relative clauses, is seen to account for aspects of their behaviour which a Sentence / Noun Phrase grammar for them (of the sort proposed by Chattopadhyay (l976a, b)) cannot account for.

Further, my theory entails a generalization which I call Simplexity. I show at the end of the chapter that Simplexity covers constituent questions as well.

8.1 Intransitive and transitive Particles

Ample exemplification has been given of the general fact that Bangla puts the head of its constructions after, rather than before, the satellite(s). It will thus come as no surprise that, when a Particle governs a sister, it follows this sister. Thus, *tel chaRa* "oil without" corresponds to *without oil*. My use of the term Particle, to avoid constantly distinguishing between Prepositions as in English and Postpositions as in Bangla, is due to Emonds (1976). This is no innocent change of terminology. It involves the claim, explicitly made by Emonds for English, that Prepositions/Postpositions are transitive Particles, and that what other frameworks call Particles are, so to speak, intransitive Prepositions/Postpositions. Like Emonds' claim for English, my corresponding claim for Bangla is based on the occurrence of the same element sometimes with and sometimes without a satellite, and on the application of Occam's razor.

(1) ram pOre jabe
"Ram after will-go"
'Ram will go later'

(2) ram SEmer pOre jabe
"Ram Shyam-GEN after will-go"
'Ram will go after Shyam'

(3) SOngge boygulo elo
"with books came"
'The books were brought along'

(4) ramer SOngge boygulo elo
"Ram-GEN with books came"
'With Ram came the books'

(5) SOngge SOngge TEksi haWa hoYe gElo
"with with taxi air become-and went"
'Immediately all taxis disappeared'

(6) chobi SeS hOWar SOngge SOngge TEksi haWa hoYe gElo
"picture finished being-GEN with with taxi air become-and went"
'The moment the film ended, all taxis disappeared'

It would be absurd to describe the odd-numbered examples above in terms of a class of Postpositions and the even-numbered ones in terms of a class of Particles distinct from but systematically homonymous and synonymous with these Postpositions. I therefore conflate the two sets into a single class of Particles. Some of these Particles occur both with and without an overt satellite, as shown. Some Particles always occur with an overt satellite:

(7) nirdharito pronali onuSare kaj korbo
"decided method according work will-do"
'I'll work according to the method decided on'

(8) *onuSare kaj korbo
"according work will-do"
[intended reading: 'I'll work accordingly']

(9) eTar bOdole oTa naW
"this-GEN instead that take"
'Take this instead of that'

(10) *bOdole oTa naW
"instead that take"
[intended reading: 'Take that instead']

Other Particles always occur without an overt satellite:

(11) aste kaj kOro
"slowly work do"
'Work slowly'

(12) tOle tOle cOkranto korche
"below below conspiracy is-doing"
'They are secretly plotting'

The examples just given will raise eyebrows; some readers will ask what line I draw between AP and PP, since words like *aste* 'slowly' seem to be Adverbs (if this category is distinct from Adjectives) or Adjectives (if not).

Any answer to such a question must take cognizance of the fact that Bangla, like German, does not derive Adverbs from Adjectives in its system of derivational morphology. There is nothing like the English *-ly* suffix. Rather, the same form – which we here call an Adjective without necessarily claiming that this category stands in opposition to 'Adverb' as a completely distinct category – is used for the function of modifying a noun as is used for the function of modifying a non-noun. *Sundor gan* and *o Sundor gaY* in Bangla correspond to German *schöne Lied* and *sie/er singt schön* rather than to English *beautiful song* and *(s)he sings beautifully*[10]. To the regret of some, certain dialects of English are changing in the direction of Bangla and German in this respect.

Consequently, if a Bangla form occurs only in modification of non-nouns, one is tempted to call it something other than an Adjective. There are at least three options with a word like *aste*, which means only 'slowly' and never 'slow' (hence the ill-formedness of **aste prokriya* 'a slow process'). (A) One may regard *aste* as an Adjective with a limited distribution, lexically marked as such. (B) One may regard *aste* as an intransitive Particle, as I do, and lexically mark it as such. (C) One may regard *aste* as some third thing – an Adverb, say – and set up a class of Adverbs which have the property that, like *aste*, they never modify nouns. The third option would constitute an anti-austerity measure with nothing to recommend it, given the availability of two options which are more austere but attain the same goals. (A) and (B) differ only in that they distribute lexical idiosyncrasy differently. I have no particularly strong commitment to option (B), which I have selected, but I will give my reasons for tentatively preferring it.

The distribution of data in (7)-(10) resembles the facts about verbs like the modal *lag* (as in *Sila haSte laglo* 'Shila started-and-kept-on laughing') which lack 'solitary' occurrences. And the distribution in (l)-(6) resembles the facts about verbs like *kha* (as in *Sila khabe* 'Shila will eat' and *Sila mach khabe* 'Shila will eat fish'), which appear with or without an overt object. In other words, straightforward pattern expectations lead one to expect there to be

[10]Not that the character of English adverbs in *-ly* is itself clear. Here is a slightly adventurous proposal, not made earlier, to the best of my knowledge, in the literature of English linguistics. Its prepositional character notwithstanding, English does have some Postpositions like *aside* and *notwithstanding*. Perhaps *ly* is a Postposition which always governs AP; the first constituent of *very quickly* is then *very quick*; cf. *for real*, where the Preposition *for* governs an AP *real*.

Particles which resemble such verbs as *ghOT* ‘happen, take place’, which never take an object (or a VP complement). This expectation is met only if one selects (B) and not (A).

Acceptance of (A), on the other hand, would also have filled a similar gap, elsewhere. Some Adjectives, like *ghumonto* ‘sleeping’ (it cannot mean ‘sleepingly’), must be marked as necessarily modifying nouns and no non-nouns, so one would expect there to be some which go the other way; and *aste* and its ilk seem perfectly good candidates for such a lexical marking and for membership in the category of Adjectives. What, then, is the argument in favour of (B) and against (A)? Aren’t the two situations parallel?

The difference is that, in (B), we are speaking of a PP structure and of strict subcategorization relations within it, whereas in (A) we are speaking of a loose-knit sentence structure (with no VP constituent when the verb is finite) where the relation between an Adjective and the Verb it modifies is not formalizable in terms of strict subcategorization in the lexical entry for that Adjective but must be relegated to selection. It is hard to document an impression, but I have the impression that virtually all linguists who have a view on the matter see strict subcategorization as a more ‘central’ aspect of lexical idiosyncrasy than selection, more ‘central’ in the sense of needing more attention in the grammatical description proper. It is on this feeling that I base my preference for (B) over (A).

This leaves me with an asymmetry between Adjectives marked as necessarily modifying nouns (such Adjectives exist) and Adjectives marked as necessarily modifying non-nouns (such Adjectives, I have decided, do not exist). It seems to me to be a tolerable asymmetry. Adjectives and nouns are linked both in the NP structure where they contract strict subcategorization relations and in terms of grammatical relations since the head noun is the ‘subject’ of the AP in some sense. But Adjectives and Verbs are not so closely related.

This argument in favour of choosing (B) over (A) is obviously based on empirical factors, although its thrust is methodological. First, it relies on my analysis of Bangla finite clauses as lacking a VP. Second, it arises from the fact that we have no data as yet that would help us to decide between (A) and (B). Thus, the argument is empirical after all, in the sense that it depends on the absence, so far, of crucial facts choosing between (A) and (B). Linguists who reject my argument must either argue for (A) or write a workable analysis which obviates the need to choose, e.g. an analysis in terms of categorial neutralization. Current linguistic theory, perhaps unfortunately, does not enable us to say that something belongs to an ‘archi-categoreme’ subsuming PP and AP but does not belong either to the category PP or to the category AP.

To sum up: I have postulated and tentatively justified a class of Particles some of whose members are rigidly intransitive or transitive and some of whose members appear both with and without an overt object.

8.2 Particles governing Noun Phrases and Adjective Phrases

Particles which govern NPs require the head N to be in some Case or other, quite idiosyncratically. The P *hoYe* ‘via’, for example, requires a Nominative head N, while *hoYe* ‘on behalf (of)’ requires a Genitive one.

There are only two Particles that specifically govern APs – *bhabe* and *kore*, both of which have an adverbializing effect; we gloss them as ‘-ly’. They leave the A stem unaltered. The fact that the Particle *kore* ‘apiece, each’ also governs NUM′ counts as evidence against the

natural alternative course of regarding *kore* and *bhabe* as (periphrastic) suffixes forming Adverbs from Adjectives. Stronger evidence for the same conclusion is available elsewhere: the P *bole*, preserving exactly the same meaning 'as', governs both NP and AP. Another piece of evidence comes from the construction *aste kore* "slowly PARTICLE" 'slowly', roughly synonymous with *aste*, where the word *aste* is an item which never occurs as a 'true Adjective' – it never modifies an N.

There are Particles imposing the Nominative, the Objective, the Genitive, or the Locative Case on the Nouns they govern. Some examples:

(13) ayn onuSare "law according" 'according to the law'; Kolkata theke "Calcutta[11] from" 'from Calcutta'

(14) tomaY/tomake niye "you-OBJ about" 'about you'; bimOlke chaRa "Bimal-OBJ without" 'without Bimal'

(15) kolkatar theke "Calcutta-GEN from" 'from/than Calcutta'; baRir pechone "house-GEN behind" 'behind the house'

(16) kaMdhe kore "shoulder-LOC on" 'on (one's) shoulder(s)'

8.3 The Locative-Conjunctive analysis versus Particles

Nobody will dispute the claim that most Particles originate historically from locative forms of nouns on the one hand (e.g. *onuSar-e, pechon-e*) and conjunctive forms of verbs on the other (e.g. *theke, kore* /thak-ye,kOr-ye/). But one may argue over whether the synchronic grammar of the language should recognize this fact, and, if it should, to what extent.

Lyn Ohira, Ngo Thanh Nhan, and Marian Maddern have suggested to me that Bangla (and languages that resemble it in this respect) can be said to have no class of Particles, and that a describer of Bangla should, instead of postulating a class of Particles, syntactically treat some putative Particles as locative nouns and the rest as conjunctive verbs while lexically recognizing that these items have a restricted range of occurrence and a specialized set of meanings.

A direct defence of the category P against this counterproposal can be based on Particles like *mOto, chaRa, bOrabOr, dara, maphik, boy,* and *onujayi*. Of all these forms only *chaRa* can be suspected of having anything to do with a noun or verb; it means 'except' or 'without'; the verb *chaR* means 'leave' or 'doff'; the noun *chaR* means 'material left out' (as in proof-reading); and the participle-gerund *chaRa* derived from the verb *chaR* means 'leaving, doffing, left, doffed'. It seems likely that the particle *chaRa* comes from the participle-gerund *chaRa*, historically. Chatterji (1926: 769) and Klaiman (1977: 305) discuss the historical relation between the participle-gerund *chaRa* and the particle *chaRa*. But the synchronic distribution of these two items is so different that only the most fanatical 'anti-particulist' would try to stretch the participle-gerund *chaRa*'s lexical entry to accommodate the facts about *chaRa*'s particle use; and this exercise in stretching, if carried out, would be wasted on only one

[11] Today, of course, the city is called *Kolkata* in English; but it would be anachronistic of us to emend *Calcutta* in these glosses.

participle-gerund in the whole language – recall that other verb-derived particles have the appearance of a conjunctive in /ye/, not that of a participle-gerund in /Wa/[12]. The remaining forms *mOto* 'like', *bOrabOr* 'alongside', *dara* 'by (means of)', *maph-ik* 'according to', *boy* 'but' (the Particle *but*, as in 'nothing but'), and *onujayi* 'according to' are impeccably pure particles. Examples of their use follow.

(17) gacher mOto (*moto* in many dialects)
"tree-GEN like"
'like a tree'

(18) raSbihari bOrabOr
"Rashbehari along"
'along Rashbehari Avenue'

(19) jhunur dara
"Jhunu-GEN by"
'by Jhunu'

(20) nijer ruci maphik/onujayi
"self-GEN taste according"
'according to one's own taste'

(21) du din boy to nOY
"two day but COMP isn't"
'(Life) is but a two-day (game)'

Another consideration which weighs against the Ohira-Nhan-Maddern suggestion relates to words like *Samne* 'in front (of)' and *pOrne ~ pOrone* 'on' (as in 'They have jackets on') which would, on the Locative-Conjunctive hypothesis, have to be analysed as Locative forms of the nouns *Samn* and *pOrn ~ pOron* meaning something like 'front' and 'wear', respectively. But the phonotactics of the language forbids a final cluster with two nasals and marks a final *rn* as characteristic of the English-derived stratum of the lexicon, to which stratum this word *pOrn* does not belong (its variant /pOron/ conforms to standard Bangla phonotactics, but the fact that the P is overwhelmingly often pronounced /pOrne/ rather than /pOrone/ needs to be kept in view). To avoid this difficulty, one might enter the locative form as such into the lexicon and postulate a defective declension. But such measures would accord indirect recognition to the difficulty of maintaining the thesis that there is no class of Particles.

In a far less difficult case – the possibility of saddling English with a new preposition *a* which would occur in 'phrases' like *atop, asleep, away, amiss, awry, amidst* (to be analyzed as $[_{pp}[_{p}a][_{NP}top]]$ etc.) – despite the relative ease with which one can break into the synchronic isolation of this group of forms, the ease with which one can set up the preposition *a* without confronting phonotactic or other arguments against such a proposal, generative grammarians have been remarkably reticent to put forward such an analysis. What deters them is probably a

[12] In certain dialects of Bangla there are two exceptions to this statement: *hote* 'from' and *cayte* 'than'; these words derive, respectively, from the verbs *hO* 'be(-come)' and *ca* 'look, want'; they are infinitival in appearance.

feeling that such highly unproductive patterns deserve no recognition in a syntax, since there is no psychological reality to them even if the pattern is formally neat. The same feeling should deter us from trying to dissolve the synchronic category of Particles in Bangla into its historical fractions, especially since this effort is fraught with operational difficulties.

To tilt the balance of arguments even further away from the Conjunctive-Locative hypothesis, I point out that the latter has no principled protection against the occurrence of one Emphasizer each for what their analysis would treat as two distinct NP nodes in, say, *[$_{NP}$ [$_{NP}$ [$_{NP}$ [$_{NP}$ naToker][$_{EMP}$ i]][$_{N}$ jonne]][$_{EMP}$ o] "play-GEN-EMP for-EMP" 'also for the *play*', which is ill-formed; cf. the well-formed [$_{NP}$ [$_{NP}$ [$_{NP}$ [$_{NP}$ Suggriber][$_{EMP}$ i]][$_{N}$ dada]][$_{EMP}$ o]] "Sugriva-CEN-EMP older-brother-EMP" 'also the older brother of *Sugriva*'.[13] Since the Particle hypothesis assigns to the ill-formed but not to the well-formed string the structure [$_{PP}$[$_{NP}$[$_{NP}$…er][$_{EMP}$ i]][$_{P}$…]][$_{EMP}$ o]], this hypothesis can resort to the principle, established during our discussion of emphasizers, that a non-S-daughter NP can have an EMP of its own only if it is an NP-daughter, which the embedded NP here is not – it is, rather, a PP-daughter. The Conjunctive-Locative hypothesis cannot explain the ill-formedness of *naTokeri jonneo* in terms of the structure of the string, but must resort to separate lexical specifications for all 'postpositionally used nouns' like *jonn-*, stating that an NP sister to any of these nouns must not get an EMP which would otherwise be allowed. Such a stipulation would be a trifle odd even if made only once in the lexicon, let alone sixty times (or whatever the number turns out to be).

On the other hand, it does look as though the Locative-Conjunctive hypothesis, unlike the Particle hypothesis as I have formulated it so far, can economically describe the fact that an NP in the Genitive governed by a 'noun-derived' Particle may take an Emphasizer (not one of its own but one 'stolen' from the PP as a whole) while other P-governed NPs cannot do this. Thus, compare *naTokeri jonne* 'for the play', well-formed, with *kolkataro ceYe* 'than even Calcutta', ill-formed (the P *ceYe* is V-derived, although the NP is in the Genitive), and *ayno onuSare* 'according even to the law', ill-formed (the NP is in the Nominative, although the Particle *onuSare* is N-derived). This 'test' makes the otherwise un-N-like *mOto* and *dara* of (17) and (19) appear to be N-derived (as they historically are); *gacheri mOto* and *jhunuri dara* are well-formed. Can my proposal cope with these facts?

My solution to the problem is to assume that these strings have the structure [$_{PP}$ NP N] at the same time as [$_{PP}$ NP P]; this is just one step removed from the 'archi-categoreme' idea which I shrank from in section 8.1. I admit that it is a little odd to have a [$_{PP}$ NP N] structure, but it allows me to apply my rule of EMP Preposing to derive the forms under discussion without violating the No New Satellite Condition. Furthermore, this modification of my proposal bends enough toward the Ohira-Nhan-Maddern counterproposal to accommodate the only fact which pleads for their Locative-Conjunctive hypothesis, but does not bend too far and thus avoids the problems.

For example, since I am saying that words like *Samne* and *pOrne* are at the same time N and P – and assuming a definition of Head which requires that the Head of a PP be only a P, not simultaneously a P and something else (say, an N) – I am recognizing them as words with no

[13] In 2020 it is clear to me, and surely to the reader, that this argument in its present form gets the analysis of /Sugriberi dadao/ wrong. We would assume, today, that such a string must have moved /Sugriberi/ out of the nominal into a clausal position. However, the argument can be recast on the basis of our current assumption – the point would now be the impossibility of moving /naTokeri/ out of the position governed by the P /jonne/.

complete allegiance to either category, as words which in their Noun role (or semi-role) need not be identified as Locative and thus need not be segmented *Samn-e* and *pOrn-e* in violation of phonotactics. This is not a trick. It is morphologically necessary to regard *mOto* and *dara* (which are in the same boat, as explained above) as non-Locative. Furthermore, since I am not saying that the entire structure is an NP, I do not lose my explanation of why the NP P sequence (though its P may simultaneously be an N) must have at most one EMP, since I can still apply my principle that only an NP dominating an NP can have two Emphasizers.

Since I am making crucial use of the fact that the mother node continues to belong to the PP category even when its contents are regarded as NP N instead of NP P, the move I have made does not force me to confront the problems of the full-fledged archi-categoreme idea. I am still dealing with fairly conventional notions, not with a neither-N-nor-P-nor-quite-both bundle of features manifested at all bar levels from N/P to NP/PP; although I have undermined somewhat the distinction between N and P for some words, the distinction between NP and PP is still rigid.

I take it now as established that apparent Particles are indeed actual Particles in Bangla, whatever else they may also be. The following considerations will, for the sake of completeness of description, lend this conclusion further support which is by now unnecessary.

When a PP functions as modifier to an N, as a constituent of an NP, its head PP (to use the term **head** loosely, i.e. paying no attention to whether the head is a through-and-through Head in the sense clarified above), if historically a locative, appears in the genitive instead, under certain conditions. Alongside *aguner kache boSlam* 'I sat down near the fire' with the 'locative' *kache* we have *aguner kacher ceYarTa* 'the chair near the fire' with *kacher*, which appears to be the genitive of the noun *kach*, just as *kache* appears to be its locative. This effect is not constant for a given Particle. The 'locative' form appears if the N modified by the PP is an action noun: *aguner kache nac* 'a dance near the fire'. And there are particles which, though apparently locative nouns, do not show the alternation. Grammatical *oMr jonne ceYar* 'a chair for him/her' and *oMr jonne nac* 'a dance for him/her' contrast with ungrammatical *jonner*. Particles which do not have the appearance of a locative noun, though they may be nouns for the purpose of the emphasizer 'test' mentioned above, remain unaltered: *SinghaSoner mOto ceYar* 'a chair like a throne' and *moYurer mOto nac* 'a dance like a peacock('s dance)' use the same form *mOto* 'like'.

Particles which show the alternation, like *kache ~ kacher* 'near', can also bear the genitive ending with an action nominal, with a slight alteration of meaning. Compare *aguner kacher nacTa* 'the dance near the fire', which singles out that dance, in contrast to, say, the dance near the window, and *aguner kache nac aguner kache ayskrim khaWar ceYe nirapOd* 'a dance near the fire is safer than eating ice-cream near the fire'. It is useful to note that a PP like *aguner kache* and an NP like *gOrom ghOr* 'a hot/warm room' behave identically in this respect: *gOrom ghOrer nacTa* 'the dance in the hot room', *gOrom ghOre nac gOrom ghOre ayskrim khaWar ceYe nirapOd* 'a dance in the hot room is safer than eating ice-cream in the hot room', *gOrom ghOrer ceYarTa* 'the chair in the hot room', ill-formed (as a constituent – not, of course, as a string within a sentence) *gOrom ghOre ceYarTa*.

Only a detailed study of the grammar of Case will reveal the secrets behind these phenomena. I will not attempt that task here. Assuming that there is some account of what is happening for *gOrom ghOr* and NPs like it, I ask whether such an account will carry over to PPs like *aguner kache*, assuming, as in the discussion of emphasizer facts above, that a P in such PPs is also marked N without losing its P character.

The answer seems to be – yes it will, but with some difficulty. The difficulty arises from the fact that 'noun/particles' which look as though they are not in the locative, like *mOto* and *dara*, and at least one 'noun/particle' which has a locative look, *jonne*, never occur with the genitive *-r* marking. The solution to this problem may lie in a notion of 'nouns of location' which excludes such words as *mOto* 'like', *dara* 'by (means of)', *jonne* 'for' on grounds which have nothing to do with morphology (although morphology may well be an independent ground for exclusion – explaining the exclusion of *paSapaSi* 'beside' for speakers who do exclude it: it is marginal for me) and thus make it unsurprising that the class thus excluded should have both locative (*jonne*) and non-locative (*mOto, dara*) members. (The word *pOrne ~ pOrone* 'on' needs to be considered a location noun.) The generalization governing the alternation of genitive with locative N forms in the context $[_{NP}[NP_{gen}$ ___] ... N], then, should be based on the explicit assumption that the N within ___ must at least be a location noun, whether or not it is also a particle.

At this point we seem unfortunately to have succumbed to a view which forces us back into the undesirable segmentations shown in *Samn-e* and *pOrn-e*. Or have we? Perhaps the genitive forms *Samner* and *pOrner* are really *Samne-r* and *pOrne-r*, and not the dreaded *Samn-er*, *pOrn-er*. If noun-particles as well as plain nouns take the genitive, that would take care of this problem and also predict, correctly, that one should be able to say *aster bOdole taRataRi* "slowly-GEN instead fast" 'fast instead of slowly', where the intransitive particle *aste*, which no one has sought to derive from any living noun but which may well be synchronically a noun/particle, shows up in the Genitive without the implication that *aste* must itself be seen as a locative. We can surely extend this approach to *Samne* and *pOrne*.

Another set of data to consider in this connection is the set that occurs in the environment ___ *theke*: *dur theke* 'from afar' (cf. *dure* 'far'), *kach theke* 'from near' (cf. *kache* 'near'), but *Samne theke* 'from in front', with its *-e* intact for reasons which are shown to be other than phonotactic by the existence of the form *age theke* 'from before = beforehand' (cf. *age* 'before'), which could phonotactically just as well have been **ag theke* (*ag*, though not an actual word, is a possible word in Bangla phonotactics). This data set seems to me to indicate that, of the locative-looking noun/particles, some are more 'really locative' than others in a way whose complexity can only be expressed if we have an independent category of Particles whose members, in their synchronic properties, have attained varying degrees of distance from the morphological allegiance they used to have to another category centuries ago.

8.4 Particles governing Sentences

Having established the existence and described some properties of the category P, we now look at phrases consisting of an S followed by a P governing it.

(22) uni baRi korechen bole ami Suni ni
"(s)he-HON house has-done as I hear not-PERF"
'I have not heard (anything to the effect that) (s)he has built a house'
(or 'has had a house built', a detail left implicit from this point on)

(23) uni baRi korechen bole oMr khub dena hoYeche
"(s)he-HON house has-done as (s)he-HON-GEN very debt has-happened"

'(S)he has fallen into great debt because (s)he has built a house'

That is all there is to exemplify: *bole* is the only S-governing P, and it has only these two senses – a purportive 'that' and a 'because'. The purportive nuance of the first sense is so marked that no epistemic operator can override it – contrary to the expectation of a speaker of English that *jan* 'know' should be able to:

(24) uni baRi korechen bole ami jantam na
"(s)he-HON house has done as I knew not"
'I didn't have the understanding that (s)he had built a house'

The peculiar English rendering is used to indicate that the sense is akin to 'if (s)he had built a house, I didn't know it'. To avoid misleading the reader about the properties of *jan* I should point out that it is very different from *know* even in sentences (and these call for a different order of clauses) where *bole* is absent.

(25) ami *jantam* maSi ratre phirben (*stressed*)
"I knew mother's-sister night-LOC will-return"
'I knew (my) aunt would return in the evening'

(26) *ami* jantam maSi ratre phirben (*stressed*)
"I knew mother's-sister night-LOC will-return"
'My understanding was that (my) aunt would return in the evening'

If anything, (26) insinuates the entailment 'but I was wrong', and does this so strongly that no assertion to that effect need be explicitly added. In other words, (26) and (25) are antonymous. It is unreasonable, then, to expect the epistemic force (which turns out not to exist, in Bangla) of the verb *jan* to override the hedginess of *bole* in (24) – even *jan* on its own, without *bole*'s company, is perfectly compatible with hedging. No word in Bangla has per se the property of *know* which causes an assertion like *I knew that he would leave* in English to commit the asserter to the truth of the proposition *He left*; only the joint effect of *jan* and the stress on *jantam* in (25) can mimic *know* in this respect.

Returning to the behaviour of *bole*, I now point out that structures like (24) are especially 'hedgy', so much so that, if you superimpose on (24) the truth-stressing intonation shown for (25), the result is bad:

(27) *uni baRi korechen bole ami *jantam*
"(s)he-HON house has-done as I knew"
intended reading: 'I knew (s)he had built a house'

This property of *bole* permits a PP with *bole* to contain the hedgiest clause type of all – questions; cf. (28). Consider, in contrast, the construction without *bole*, in (29) and (30).

(28) ke baRi koreche bole tumi bolechile?
"who? house has-done as you said"
'Who did you say had built a house?'

(29) tumi bolechile ke baRi koreche
"you said who? house has-done"
'You told (me) who had built a house'

(30) tumi ki bolechile ke baRi koreche?
"you COMP said who? house has-done"
'Did you tell (me) who had built a house?'

Using the construction of (29) one cannot say what (28) says – avoid commitment to the claim that someone had built a house, avoid speaking of the addressee's revealing the identity of the person who had actually built a house. The only way in which this construction can be used to pose a question is to question the matrix sentence, as in (30) which asks whether the addressee revealed the identity of the person concerned but cannot, within the confines of the construction, ask who this person is. To give 'matrix' effect to the 'embedded' question, so to speak, one resorts to the construction in (28), with *bole*.

As with questions, so also with relative clauses, *bole* 'passes up' the force of the relative word to the next-higher S.

(31) je_i baRi koreche bole tumi bolechile Se_i rOmeS nOY
"who house has-done as you said he Romesh isn't"
'The person who you said had built a house isn't Romesh'

Again, a corresponding sentence using the construction without *bole* prevents the relative from being linked with an antecedent in the matrix S: (32), with *je baRi koreche* embedded and the sequences *tumi bolechile* and *Se rOmeS nOY* higher up in the tree (and with the expectation that *Se* will play antecedent to *je* as in (31)), is ill-formed and uninterpretable. The only way in which one can interpret this string is by assigning it the structure of (33), where *tumi bolechile* functions as reportive matrix to the entire sequence *je baRi koreche Se rOmeS nOY*, and where *Se* and *je*, both being in the embedded structure, can now be linked anaphorically, yielding an interpretation quite different from that of (31).

(32) [[tumi bolechile [je_i baRi koreche]] Se_i rOmeS nOY]
"you said who house has-done he Romesh isn't"
intended reading: 'The person who you said had built a house isn't Romesh'

(33) [tumi bolechile [je_i baRi koreche Se_i rOmeS nOY]]
"you said who house has-done he Romesh isn't"
'You said the person who had built a house wasn't Romesh'

This ability of *bole* to 'pass up' questions and relatives to the next higher S makes it functionally rather similar to the English complementizer *that*, which performs the same function in sentences like *Which house did you say that he had built?* and *This is the house which you said that he had built*; the difference is just that in English the question phrase or relative phrase actually travels up past the *that*, whereas in Bangla it is only the effect of the question phrase or relative phrase which 'travels' upward.

In view of this similarity some readers are bound to ask me to justify my decision not to treat *bole* as a COMP; another way to put the query is 'should S + *bole* sequences be called Complement Clauses?' The answer is that, unlike a Complementizer, *bole* (a) can co-occur with a Complementizer in the same simplex S′ and (b) never undergoes COMP Preposing.

(34) lusi je phel korbe bole Tunu bhebechilo raka ta jane ki?
"Lucy COMP fail will-do as Tunu thought Raka that-fact knows COMP"
'Does Raka know that Tunu thought Lucy would fail?'

(35) *lusi bole je phel korbe Tunu bhebechilo raka ta jane ki?
"Lucy as COMP fail will-do Tunu thought Raka that-fact knows COMP"

Therefore, *bole* is not a Complementizer, and the S + *bole* construction cannot be reasonably called a Complement Clause, despite the similarity of semantic function to Complement Clauses in languages like English.

Klaiman (1977) states that *bole* is a COMP. She gives no arguments for her view, and I have just argued against it. Udaya Narayana Singh, in an as yet unpublished paper in Bangla which I have read but at the moment of writing do not have access to, states that *bole* is a noun that takes an S complement; he is apparently led to this view by his decision not to regard *bole* as a Complementizer. However, *bole* shows no noun-like behaviour with respect to declension or derivation or NP structure, and the view that an S + *bole* structure is an NP puts the *je-Se* relation in (31) in conflict with general principles of sequent-relative binding (within the CS research tradition, the relevant principle is the Complex NP Constraint; for the story in my framework, see the rest of chapter 8). Both Miriam Klaiman and Udaya Narayana Singh, besides, will have difficulty reconciling their views with the fact that *bole*, like bona fide particles, governs NP and AP.

8.5 The syntax of Relative Clauses

I deliberately hedged about the constituent structure of (32) and (33) above, since I have yet to answer some fundamental questions about the syntactic relation between a relative clause in Bangla and its antecedent phrase. These questions arise because of sentences like (36), which behave unlike relative-clause-containing sentences in languages like English and French – where everybody assumes relative clauses and antecedent phrases to be syntactically linked.

(36) je chele je puroSkar caY Sey chele Sey puroSkar pabe
"which boy which prize wants that boy that prize will-get"
'The boy who wants it will get the prize he wants'

One cannot build for (36) a deep structure affiliating the relative clause *je chele je puroSkar caY* either to the NP *Sey chele* or to the NP *Sey puroSkar*. If one affiliates it in deep structure to *Sey puroSkar*, one is obliged to, and fails to, (i) find a well-founded rule that shall prepose the clause, yielding (36), and (ii) offer a good reason for the ill-formedness of the surface structure *Sey chele je chele je puroSkar caY Sey puroSkar pabe* "that boy which boy which prize wants that prize will-get".

If one affiliates the clause to *Sey chele* in deep structure, one still faces the challenge of justifying the claim that *je chele je puroSkar caY Sey chele* is a constituent at any level of structure. This sequence shows no unitary grammatical behaviour of the sort one would expect of a constituent. We conclude that the structure of (36) does not have a constituent which includes the relative clause and one NP and nothing else.

We must now further ask whether the relative clause is a sister clause to *Sey chele Sey puroSkar pabe* – the Coordinate Clause Hypothesis, CCH – or a subordinate clause which is a sister to the pieces of the sequence after it (e.g. a sister to *Sey chele,* to *Sey puroSkar,* etc.) but not to the sequence as a whole since the latter is not a constituent – the Subordinate Clause Hypothesis, SCH.

Lyn Ohira has pointed out to me the following argument for CCH. Bangla permits parataxis, coordination by mere juxtaposition. Thus, its grammar must allow S to dominate S S. Besides, some clauses, exclamations, occur as root sentences containing J-Words. Thus, lexical insertion must introduce J-Words freely. Therefore Bangla grammar will as a matter of course generate strings like (36) with that S S structure unless you make a special effort to prevent this. If there is something wrong with the CCH structure for (36), it needs to be pointed out and a reasonable procedure given to block the derivation of an S S structure for (36). Otherwise one should define one's semantics for relatives on the S S structure, instead of concocting a special relative syntax with extra apparatus and inventing reasons to rule out the coordinate structure in (36). Occam's razor places the burden of proof on the proponent of SCH, who will need to dejustify the S S structure for (36) and to justify the extra apparatus needed for SCH.

As a proponent of SCH I accept the burden of proof. My first argument for SCH is that a bona fide coordinate S S structure exhibits optional and occasionally obligatory terminal contour intonation at clause break. But (36) does not do this. Therefore, there is no clause break; *Sey chele Sey puroSkar pabe* is not a clause; SCH is correct.

My second argument rests on sentence (37), which contains the relative clause (38).

(37) amra oke o je dokan theke je jiniS ceYeche Sey dokan theke Sey jiniS ene debo
"we him/her (s)he which shop from which thing wants that shop from that thing bring- and will-give"
'We'll bring him/her the thing(s) (s)he wants from the shop(s) (s)he wants it/them from'

(38) o je dokan theke je jiniS ceYeche
"(s)he which shop from which thing wants"

The relative clause here occurs within, not on the periphery of, its matrix sentence. It follows *amra* and *oke* and precedes the sequence *Sey dokan theke Sey jiniS ene debo*, whose main verb, *debo*, is in the first person, agreeing with the subject *amra* of that verb. Thus, one must regard (38) as an immediate constituent of (37) on a par with *amra, oke*, *Sey dokan theke, Sey jiniS*, and *ene debo*. Therefore (38) cannot even in deep structure be one of two coordinate clauses. No well-founded transformation would take us from such a deep structure to its putative surface reflex.

A third argument against CCH stems from the ability of each member of a coordinate structure to substitute, in a structuralist sense, for the coordinate structure, since 'coordinate' presupposes 'endocentric' and entails equal 'centricity'. But neither (38) nor the relative clause in (36) can stand on its own as well-formed. For obscure reasons within the unstudied grammar

of Bangla exclamations, (39) and (40), unlike (41), are unacceptable exclamations.

(39) je chele je puroSkar caY!
"which boy which prize wants"

(40) o je dokan theke je jiniS ceYeche!
"(s)he which shop from which thing wants"

(41) mitul aj ja ranna koreche!
"Mitul today what cooking has-done"
'Oh, the cooking that Mitul has done today!'

Only a small number of exclamatory formulas containing J-Words seem to exist, anything outside this range being distinctly ill-formed.

This leads into a fourth argument to the same effect. The few cases where a simplex root S may contain a J-Word are exclamations and, as such, do not coordinate, and have other special features, particularly in phonology. These features distinguish them from the many (infinitely many, as is usual in syntax) J-Word-containing clauses which occur as relative clauses. If you maintain SCH, then, by adding other mechanisms to handle exclamations with J-Words in them, you can begin to account for this exact correlation between form and function. But if you maintain CCH, I think you are committed to a theory which equally elegantly and productively generates an infinite set of J-Word-containing exclamations, contrary to fact, alongside the infinitely many relative clauses which do exist.

Having offered positive evidence for SCH, I now respond to Lyn Ohira's demand (that I come up with a reasonable way to prevent the derivation of CCH structures for relatives by independently given machinery of the grammar) by appealing to Schachter's (1977) reformulation of the Coordinate Structure Constraint. His reformulation requires that coordinated items be similar in fundamental grammatical properties. A relative clause thus cannot coordinate with anything other than a relative clause. Therefore, the normal mechanisms do not generate CCH-structured relative clauses. Occam's razor no longer supports CCH.

Taking the issue of SCH versus CCH to be settled, I turn now to other sorts of relative clauses, for which a slightly stronger theory than SCH seems indicated. There is some evidence that a relative clause containing only one relative phrase (instead of two, as in (36) and (37)) can team up with the antecedent to form a constituent. Thus, we get fragmentary NP utterances like *dalal-Talal ebong jara biplOber Sotru tara* "agents-etc. and who-PL revolution-GEN enemy they" 'agents-etc. and those who are hostile to the revolution'. Such NPs also have a unitary semantic interpretation; one might view the interpretation of the specifically relative-laden conjunct in terms of Chomsky's set intersection proposal discussed later in this chapter. One could generate relative-laden NPs by using analyses along the lines of those available for English or French relative clauses (with the clause to the left rather than, as in English or French, to the right of its antecedent). Alternatively, one could have a transformation derive such NPs from deep structures such as (42), which the grammar also needs:

(42) $[_S$ $[_S$ jara biplOber Sotru $[_{NP}$ tara] ghapTi mere thakbe]
"who-PL revolution-GEN enemy they low lying will-stay"
'Those who are hostile to the revolution will lie low'

(43) surface structure: [$_S$ [$_{NP}$ [$_S$ jara biplOber Sotru][$_{NP}$tara]] ghapTi…]

Occam's razor dictates that (42) be the deep structure and (43) the surface structure. I will call this decision of mine the Extended Subordinate Clause Hypothesis (ESCH).

By touching on this matter I have stepped on territory already covered by Jayanti Chattopadhyay (1976a, b), who proposes the structure in (43) as the deep structure for all relative clauses in Bangla; I will call this proposal the Sentence-NP hypothesis (SNP). Before proceeding further, I need to compare her analysis with mine, for the two grammars differ in basic respects. Her analysis inserts the relative-marking J element and the antecedent-marking T element by means of insertion transformations, while I base-generate these elements. Many of the differences of detail between the grammars follow from the fact that Chattopadhyay's work belongs to the generative semantics tradition of research and uses such powerful and homogeneous mechanisms as global derivational constraints, while my work belongs to the interpretive semantics tradition and uses much more severely constrained and differentiated sets of mechanisms.

Chattopadhyay's SNP grammar introduces all relative clauses by means of the PS rule S → S NP, requires that the S contain an occurrence of NP_i identical to the NP_i sister of the S, and uses an obligatory transformation introducing simultaneously an occurrence of the Relative DET *je* into the lower NP_j and an occurrence of the antecedent-marking Sequent DET *Se* into the higher NP_j. Although she does not formalize this transformation, her remarks and examples (Chattopadhyay 1976a: 32-33) permit the inference that she would formalize it roughly as follows, taking + to mean daughter-adjunction.

(44) Relative and Sequent Marker Placement (obligatory)
S.D.: X – [$_{NP}$[$_S$Y – NP – Z] – NP] – W
S.C.: 1 2 3 4 5 6 ➔1, 2, *je*+3, 4, *Se*+5, 6
Condition: 3=5

In contrast to SNP, with which (44) is associated, ESCH assumes Bangla relative clauses to be unaffiliated to any NP in deep structure, introduces these relative clauses by means of a PS rule of the form S → (S′) ... NP ... (S′) ... (S′) ... V, does not require that any relative clause contain a replica of its antecedent NP, and base-generates both the relative marking(s) in a relative and any sequent marking(s) in its matrix clause, leaving it up to semantic machinery (introduced later in this chapter) to match relative phrases up with antecedents.

In comparing ESCH with SNP, I will hold constant the fact that Chattopadhyay and I both exclude non-restrictive relatives (also known as appositive relatives), whose properties are poorly understood, and the slightly more controversial exclusion (also made by both of us) of what I will call Right Relatives in the discussion in section 8.6.

One fact which chooses between SNP and ESCH is that only ESCH allows (45).

(45) jara ayuber jonggi nitir bipokkhe chilo taMr pOton tader spOrSo kOre ni
"who Ayyub-GEN militaristic policy-GEN against were his fall them affect does not-PERF"
'Ayyub's fall didn't affect those who opposed his militaristic policies'

Under ESCH, (45) can be simply base-generated with *jara ayuber jonggi nitir bipokkhe chilo* as an S′ daughter of the root S. Since the antecedent *tader* 'them' is not contiguous to the relative S′, the option of reanalytically adjoining the S′ to the antecedent NP is unavailable, so that the deep structure remains intact, and the derivation proceeds without trouble. Under SNP, on the other hand, one must have an SNP deep structure, which means that the relative S must initially sit next to its antecedent *tader*. But *tader*, being the object of the matrix verb, normally occurs after the subject, a fact which makes (46) dubious as a deep structure and leaves (47) as the only plausible candidate consistent with SNP.

(46) jara ayuber jonggi nitir bipokkhe chilo tader taMr pOton spOrSo kOre ni

(47) taMr pOton jara ayuber jonggi nitir bipokkhe chilo tader spOrSo kOre ni

However, SNP cannot derive (45) from (47), despite its need to. Not only has no S-preposing extraposition rule been proposed or motivated for Bangla, but Baltin (1978) argues strongly that natural language excludes such rules in principle. In this particular case, even if you reject Baltin's view, a derivation from (47) would be ad hoc in that the required movement rule would lack independent motivation. Thus, ESCH can, and SNP cannot, handle (45).

My second argument against SNP is based on (36) and has a general and a specific form. The general form demolishes the SNP idea as such. It goes as follows. The sequence *je chele je puroSkar caY Sey chele* in (36) is not an NP, nor is the non-sequence *je chele je puroSkar caY — Sey puroSkar*; therefore the relative clause *je chele je puroSkar caY* cannot, either in deep or in surface structure, be affiliated to either of the NPs *Sey chele* and *Sey puroSkar*; hence ESCH and not SNP. The specific form of this argument targets Chattopadhyay's version of SNP. In order to generate (36), she would have to begin with (48) and apply (44), which would yield (49), a grammatical but different structure (it shows the boy/s in question getting some prize/s – not the particular prize/s they want).

(48) $[_S\,[_{NP}\,[_S$ chele puroSkar caY$][_{\ NP}$ chele$]]$ puroSkar pabe$]$

(49) je chele puroSkar caY Sey chele puroSkar pabe
"which boy prize wants that boy prize will-get"
'The boy who wants a prize will get a prize'

Alternatively, Chattopadhyay could try beginning with a deep structure like (50), leading her derivation to the outcome (51), which is ungrammatical and which, furthermore, conveys a meaning different from (36).

(50) $[_S$ chele $[_{NP}\,[_S$ chele puroSkar caY$][_{NP}$ puroSkar$]]$ pabe$]$

(51) *chele chele je puroSkar caY Sey puroSkar pabe
"boy boy which prize wants that prize will-get"

A proponent of (44) might argue that one could simply change (44) into (44′) in order to handle (36), without affecting the spirit of the proposal. The modified rule can apply to two or more target NP pairs.

(44′) Revised Relative and Sequent Marker Placement (obligatory)

S.D.: X – $_S$Y – NP – Z – W – NP – U

S.C.: 1 2 3 4 5 6 7 ➔ 1, 2, *je*+3, 4, 5, *Se*+6, 7

Condition: 3 = 6

It is true that (44′), applying to a deep structure of the form (48), will now generate (36) without mishap. The trouble, though, is that (44′) will also overgenerate – applying to coordinate structures to derive (53), for example, from (52).

(52) puguru-bhaSira bOrnomala caY na kintu markinra bOrnomala capacche
"Puguru-speakers alphabet want not but Americans alphabet are-imposing"
'The speakers of Puguru don't want an alphabet, but the Americans are imposing one'

(53) *puguru-bhaSira je bOrnomala caY na kintu markinra Sey bOrnomala capacche
"Puguru-speakers which alphabet want not but Americans that alphabet are-imposing"
*'The speakers of Puguru don't want the alphabet but which the Americans are imposing'

Notice that the Coordinate Structure Constraint, even under Schachter's reformulation, cannot prevent such a rule application, just as it is unable to prevent the application of rules like Gapping, for this constraint only legislates against relating the inside of a coordinate structure to its outside, and here (as in Gapping) the process goes from one wing of a coordinate conjoined structure to the other wing. Thus, a proponent of rule (44′) would have no principled means of either ensuring the survival of the well-formed structure (52) all the way to the surface (unmolested by (44′)), or blocking the generation of the ill-formed structure (53). (S)he would thus be forced to abandon (44′) as a modification of (44), and thus to acknowledge (36) as a crucial counterexample to the entire SNP idea, to the spirit and not just to the letter of rule (44).

My third argument for ESCH is as follows. Relative clauses like *protul jake Samlate pare na* in (54) exist whose antecedent is marked not by the Sequent *Se* or its allomorphs but by the Proximal Demonstrative *Emon* 'such'. (44), SNP's only route to relative constructions, fails to generate (54), while ESCH can generate and interpret (54) without adding or subtracting any basic principle.

(54) protul jake Samlate pare na Emon mOddop aj-o jOnmaY ni
"Protul whom manage can not such drunkard today-EMP is-born not-PERF"
'A drunkard such that Protul cannot manage him has yet to be born'

Fourth argument: suppose a proponent of SNP were to add an *Emon* rule alongside (44) identical to the latter except that, where (44) has *Se* 'that', the new rule would have *Emon* 'such'; and suppose this grammarian were to specify the conditions under which this rule, rather than (44), would be triggered in order to implement relativization. This would remove (54) as a counterexample against SNP, but would raise a new spectre, that of sentences like (55):

(55) –273° selsiyase-o ja drobibhuto hOY na *Emon* je paMcTa gEs ache *Sey* paMcTa gEs-i omelas grohe paWa jaY

"–273° C.-LOC-EMP what liquefied is not such which five-item gas exist that five-item gas-EMP Omelas planet-LOC getting goes"

'All five existing gases which don't liquefy even at –273° C. are available on the planet Omelas'

In (55), *Emon* is correlated to the relative *ja* while *Sey* (a variant of *Se*) is correlated to the relative *je*; no problem for ESCH. But SNP would have no way to derive the correct structure; it would derive instead at best the ill-formed sequence *…je Emon paMcTa gEs ache…* To see this, consider first the transformational cycle applying to the clause *–273° selsiyase-o ja drobibhuto hOY na Emon je paMcTa gEs ache* on the SNP analysis. The underlying structure would be $[_S\ [_{NP}\ [_S$ –273° selsiyase-o $[_{NP}$ paMcTa gEs] drobibhuto hOY na$][_{NP}$ paMcTa gEs] ache]. If, as required, we apply the new *Emon* rule, the output is $[_S\ [_{NP}\ [_S$ –273^0 selsiyase-o $[_{NP}$ je paMcTa gEs] drobibhuto hOY na$][_{NP}$ Emon paMcTa gEs]] ache], to which the further application of rules of deletion etc. (which Chattopadhyay partly sketches and partly implies) will yield *–273° selsiyase-o ja drobibhuto hOY na Emon paMcTa gEs ache*. Now move up one cycle and consider the following input to rule (44) which needs to be applied on the root S cycle: $[_S\ [_{NP}\ [_S\ [_{NPi}$ –273° selsiyase-o $[_{NP}$ ja] drobibhuto hOY na$][_{NPj}$ Emon paMcTa gEs]] ache$][_{NP}$ paMcTa gEs-i]] omelas grohe paWa jaY]. By the A-over-A condition we should relativize (i.e. treat as factor 3 in the proper analysis of the phrase-marker for (44)) NP_i rather than NP_j; this course leads to the disastrous output *je –273° selsiyase-o ja drobibhuto hOY na Emon paMcTa gEs ache Se(y) paMcTa gEs-i omelas grohe paWa jaY* "which –273° C.-LOC-EMP what liquefied is not such five-item gas exist that five-item gas-EMP Omelas planet-LOC getting goes.". Fortunately for SNP, one of the recent reformulations of the A-over-A condition would permit (44) here to be applied either to NP_j or to NP_i, although it remains problematic for SNP that nothing blocks the application of (44) to NP_i or the disastrous consequence of this application. But even if we set aside this problem and apply (44) to NP_j instead of NP_i, the output is still ill-formed: *−273° selsiyase-o ja drobibhuto hOY na je Emon paMcTa gEs ache Se(y) paMcTa gEs-i omelas grohe paWa jaY*, with the wrong order *je Emon* "which such" instead of *Emon je* "such which".

One sees, then, that the failure of SNP pointed out in the third argument is not a matter of a specific poor formulation of SNP, but pertains to the fundamental principles of SNP as an account of relative clauses.

Fifth argument: variants of sentences like (54) exist without the relative phrase – for example, (56) derived from (54), or (57-b) derived from (57-a). Presumably[14] a deletion rule derives (56) and (57-b) by removing the relative phrase. Such a deletion rule makes eminent sense in ESCH, which base-generates the relative phrase, but much less sense in SNP, which

[14]Note added in 2020: The word 'presumably' was my way, in 1979, of flagging the fact that I knew such a rule faced intractable difficulties as some of its applications violated Recoverability, a problem for which my 1979 framework had no solution (readers following my current work are aware that I have now found a semiotic solution). Note that, apart from this polemical passage, the thesis does not refer to the deletion rule, and omits it in Chapter 11.

first uses one transformation to introduce the relative element and then another transformation to remove it.

(56) protul Samlate pare na Emon mOddop aj-o jOnmaY ni
"Protul manage can not such drunkard today-EMP is-born not-PERF"
'Such a drunkard as Protul cannot manage has yet to be born'

(57)a. SiSeke ja Sona korte pare Emon drobbo kimiyar SOpno chilo
"lead-OBJ what gold make can such substance alchemy-GEN dream was"
'A substance which would turn lead into gold was alchemy's dream'

b. SiSeke Sona korte pare Emon drobbo kimiyar SOpno chilo
"lead-OBJ gold make can such substance alchemy-GEN dream was"
'Such a substance as would turn lead into gold was alchemy's dream'

Sixth argument, which is really two arguments: sentence (58-a), which exhibits 'split' control of the two relative phrases by an antecedent phrase, can be handled under SCH, but not under SNP. Even in a case where, as in (58-b), we still have a one-to-one matching and paraphrase relation between the relative phrase and the antecedent phrase – i.e. in a situation of 'pivot control' rather than 'split control' – it is possible for the relative NP and the antecedent NP to have different head nouns, again contrary to the predictions of SNP and in keeping with SCH.

(58) a. ram tOkhon jake$_i$ ja$_j$ bolechilo pOre tar$_{ij}$ SObTa-i prottahar koreche
"Ram then who-OBJ what told later that-GEN all-item-EMP retraction has-done"
(roughly) 'Ram retracted later all that he had said before'

b. je lok cukti Soy korechilo Sey manuS-i ebar rukhe daMRalo
"which man contract sign did that person-EMP now defy-and stood"
'The same man/person who signed the contract now took a defiant stand' (*lok* may, and *manuS* must, mean 'person')

Seventh argument: in the proverb *jEmon kOrmo tEmon-i phOl* "of-what-sort action of-that-sort-EMP result" 'As you sow, so shall you reap', and in more ordinary sentences which make use of *jEmon* and *tEmon*, one may regard *jEmon* and *tEmon* as A(P)s. I have found no clear arguments motivating any internal structure for these words, although presumably such segmentation (*jE-mon*, *tE-mon*) into a DET and something else is not impossible. Regardless of whether one posits internal structure, the very fact that these expressions are APs makes the SNP rule (44) incapable of handling them. But ESCH as it stands can directly accommodate the facts about *jEmon* and *tEmon*.

Having established that ESCH is better than SNP, I must now face an objection to ESCH: ESCH, by virtue of its postulation of a free adjunction rule which adjoins a relative clause to its head NP if the latter is contiguous, incorrectly generates a surface structure for the boy and prize sentence (36) in which *je chele je puroSkar caY Sey chele* "which boy which prize wants that boy" is an NP; as we know, this is a bad surface structure; ESCH also generates the correct surface structure (since the adjunction rule is optional), but the point is

that we need some way to reject the bad surface structure. My response to this objection is that a surface structure with such an NP will fail to go through the semantic component and to get a non-anomalous semantic interpretation, and that we therefore need not try to block such cases in the syntax. I am not totally happy with this response, but I have nothing better to offer, and the response is technically impeccable.

8.6 Some properties of Right Relative Clauses

J. Chattopadhyay, who proposed SNP for all Bangla relative clauses, does not draw from cases like *jara biplOber Sotru tara* 'those who are hostile to the revolution' the arguments which motivate a surface SNP analysis; she has such examples in her dissertation, but does not discuss them in detail. In contrast, another set of Bangla relatives, Right Relative Clauses, are not even exemplified in her corpus, let alone used as a basis for arguing that these clauses must have deep structures of the form [$_{NP}$...S...]. Right Relative Clauses are those which do not precede but follow their antecedent, permitting etymologically felicitous use of the term 'ANTEcedent'.

(59) a. ram Sey hire khuMjche jar kOtha tumi bolechile
"Ram that diamond is-looking-for which-GEN word you said"
'Ram is looking for that diamond which you had talked about'

b. *ram Sey hire jar kOtha tumi bolechile khuMjche
"Ram that diamond which-GEN word you said is-looking-for"

c. or doSTa ki jar jonne Sasti pacche?
"his/her fault-item what? which-GEN for punishment is-getting"
'What is the fault (s)he is being punished for?'

d. *or doSTa jar jonne Sasti pacche ki?
"his/her fault-item which-GEN for punishment is-getting what?"

e. ekhane kew kew thake jara Sukh-dukkher otit
"here someone someone live who joy-sorrow-GEN beyond"
'Some people live here who have moved beyond joy and sorrow'

f. *ekhane kew kew jara Sukh-dukkher otit thake
"here someone someone who joy-sorrow-GEN beyond live"

As these examples show, Bangla Right Relatives differ from Bangla Left Relatives in two respects. Right Relatives are always unipivotal: they contain only one relative phrase, which is pivot-controlled by an antecedent; and Right Relatives always extrapose to the end of the sentence in Bangla, whereas such extraposition is optional in English and Hindi (but cf. the contrast between *What did he do that was wrong?* and **What that was wrong did he do?*, an unexplained contrast in English).

These examples also suggest that Bangla puts the relative phrase in a Right Relative at the beginning of the Right Relative Clause. However, this positioning is optional, so that (59-g) with a different order, for instance, is acceptable, though less so than (59-h) with the clause-

initial relative phrase.

(59) g. ?o ki nite ceYechilo tumi oke ja nite daW ni?
"(s)he what? to-take wanted you him/her what to-take gave not"
'What did (s)he want to take that you didn't let him/ her take?'

h. o ki nite ceYechilo ja tumi oke nite daW ni?
"(s)he what? to-take wanted what you him/her to-take gave not"
'What did (s)he want to take that you didn't let him/ her take?'

Besides, this phenomenon is not easy to separate sharply from the general facts about emphasis-related preposing of various phrases – facts often described in terms of Bangla NPs having freedom of order.

(59) i. EkTa hirer kOtha tumi bolechile
"one-item diamond-GEN word you said"
'You mentioned a DIAMOND'

There are, then, three unclarities about Right Relative Clauses in Bangla. (A) Why is their extraposition obligatory? (B) Why do they have only one relative phrase each? (C) Do they involve Relative Phrase Preposing? Since these unclarities distinguish Right Relative Clauses from Left Relative Clauses, and since Right Relatives are a less visible part of Bangla structure than Left Relatives (both in terms of statistical preponderance and with respect to analytic priority – if you have a description of Left Relatives, which permit one or more relative phrases per relative clause, you can extend it to Right Relatives, which permit only one; but, as we have seen in connection with J. Chattopadhyay's SNP analysis, it doesn't work the other way round), I have decided to exclude Right Relative Clauses from the rest of this investigation. Except in the present section, wherever I say 'relative clause' I shall only mean 'left relative clause'.

I have no conjectures to offer regarding (A). As for (B), I suspect that something related to processing load is involved, but have not done proper psycholinguistic work and therefore will not hazard a guess. If the problem were purely internal to Bangla, I could link (B) to (A) and rely on Baltin's (1978) result that extraposed structures are interpreted via Baltin's construal schema, which, Baltin suggested to me in 1978, may explain why a Right Relative Clause should behave like a Modifier and therefore modify only one NP (since it seems impossible to modify two NPs at once – modification differs from binding in this respect). However, this cannot be the right answer, since problem (B) also exists in Hindi, and Hindi Right Relative Clauses do not always extrapose. I will devote the rest of this section to a discussion of (C). Before beginning this discussion I want to point two things out. One is that, as my use of the term extraposition and my reference to Baltin should have made clear, a Right Relative Clause and its antecedent NP do team up to form an NP constituent; there is the usual sort of evidence for the coherence of this constituent. The other thing is that, although I am making no formal claim that my analysis of Left Relative Clauses carries over to Right Relative Clauses *modulo* the special difficulties of the latter, I do actually feel that I will be able to make this claim when more is known about Right Relative Clauses. I turn now to (C).

Jim Gair, who first drew my attention to the special status of Indic Right Relatives, maintains (personal communication) that they involve Relative Phrase Preposing. In Hindi, the

language which Gair and I were speaking of, this may indeed be the case. But in Bangla forms without preposing are still grammatical; the rule is not, as it is in English and French, obligatory. Furthermore, as I have pointed out already, ordinary NPs like *EkTa hire* 'a diamond' also seem to 'prepose' under similar conditions. Perhaps there is a more general preposing rule. Perhaps NP order is, alternatively, truly free in the base to some extent. Thus, I feel that there is no warrant for postulating a Relative Phrase Preposing transformation at present. Such a rule, if postulated, would furthermore have to make crucial reference to the distinctive structure of the Right Relative construction, since Left Relative Clauses don't do it. So much complexity in the statement of a transformation is (i) suspect anyway, (ii) especially suspect when the rule is an optional one, and (iii) sufficient – in my opinion – to warrant not postulating the transformation.

It seems likely that others will argue for a rule which preposes specifically relative phrases in Right Relative Clauses. Such arguments, if they can be given, will rest on data which I have not noticed. Suppose there are such facts and that a case can be made for the preposing rule, or, equivalently, for a rule which interprets Right-Relative-Clause-initial relative phrases (and no non-relative phrases in that position) in a special way. (For the sake of simplicity of discussion, I will continue to talk about preposing alone, without mentioning this interpretive alternative every time.) If one can argue for the preposing rule, it will become necessary to ask whether the rule imperils the Complementizer Substitution Universal to the effect that only languages with COMP to the left of S have a transformation which moves Open (relative or interrogative) Phrases into COMP position. This universal, CSU, is due to C.L. Baker (1970) and J. Bresnan (1970).

Since Open Phrase Movement was the best-known unbounded movement rule around, it was apparently tacitly assumed by many linguists that CSU is about all unbounded movement rules. Thus, Kaufman (1974), who argued that Navajo had a left-hand COMP node and an unbounded postposing rule, seemed to think she had found a counterexample to CSU. That Navajo is a counterexample actually remains to be shown.

Slightly more serious is the counterexample due to Roger Epée (1976) who points out that in Duala a COMP (a question particle also occurring in alternative questions) can occur in the initial position of a constituent question and remain when an interrogative NP gets preposed into post-initial position. This is a case of a language with a left-hand COMP and an Open Phrase Preposing rule where the latter clearly does not substitute a phrase for the former. Recent literature has focused on earlier forms of English which also serve to make the point that open phrases do not take the place of complementizer words, although no one to my knowledge has pointed out the significance of the Middle and Old English data for CSU. In view of Duala, Middle English, Old English, and languages like them, CSU clearly must be restated so that it says that only a language where COMP precedes its S may have a rule which moves Open Phrases to S-initial or S′-initial position. As far as I know, no counterexamples to CSU have been offered which falsify this version of the Universal.

However, should the rule of relative phrase preposing in Bangla right relative clauses prove to be well-founded, then that will constitute a counterexample even to this version of CSU. In that case it will be necessary further to weaken CSU so that it is only about obligatory rules; the Bangla rule, even if it exists, is clearly optional.

I stress that this discussion has been purely hypothetical in that, for all we know, linguists may never find evidence that there really is a rule of relative phrase preposing in Bangla. Perhaps CSU may then be maintained in its stronger form (covering optional as well as

obligatory rules).

8.7 Chomsky's proposal for interpreting relatives

As my decision to postpone indefinitely the study of Bangla right relatives may have aroused some ire and some puzzlement, let me begin the discussion of relative clause semantics by showing that a familiar proposal by Chomsky which has been found workable for right relatives (in English) completely fails for left relatives and cannot be extended to cover the latter case without being drastically modified. Thus, left relative clauses are scientifically more interesting in the sense that it is for them that new proposals must be formulated. Analyses similar to Chomsky's are current in the research traditions of relational grammar and Montague grammar; I do not digress to consider them in detail, since the arguments I provide affect them equally. Chomsky's own analysis was formulated in dialogue with Montague grammar.

Semantic description is not being undertaken here for its own sake, but because the course of the present inquiry requires it. In chapter 7 I was obliged to reject the transformational Complementizer Transfer Hypothesis in favour of the Semantic Matching Hypothesis. The facts have forced me into semantics.

In his QFI (*Questions of form and interpretation*, in Chomsky 1977a) Chomsky offers a few sample semantic representations for relative structures. QFI is the only place in his writings that provides semantic representations for relatives. "Taking B, O, A to be (respectively) the class of books, things we ordered, and things that arrived, and taking c[*X*] to be the cardinality of *X*, we can represent the meaning of the sentences (for present purposes) as follows: (1) *the books we ordered arrived*: $B \cap O \subset A$: $c[B \cap O] \geq 2$; (2) *the book we ordered arrived*: $B \cap O \subset A$: $c[B \cap \overline{O}] = 1$ [...]; (3) *some books we ordered arrived*: $\exists K \subset B \cap O$, $K \subset A$: $c[K] \geq 2$; (4) *a book we ordered arrived*: $\exists K \subseteq B \cap O$, $K \subset A$: $c[K] = 1$. We now have the following notions: relative clause construction, which corresponds to set intersection […]" (Chomsky 1977a: 48-49).

The qualification "for present purposes" does not refer to the kernel of the proposal; it possibly refers to the fact that Chomsky is using set inclusion rather than n-place predication — set theory rather than the predicate calculus — to represent the semantics of the subject-verb relation. A later paper by Chomsky, *Conditions on rules of grammar*, says in its 38th footnote (Chomsky 1977a: 201), "On rules for relativization, see Chomsky (1975a)" – which turns out to be a reference to QFI; I offer this as positive evidence for the view that the qualification "for present purposes" does not refer to the essentials of the representations. (As negative evidence for this contention, note that neither Chomsky nor scholars associated with his approach have produced a semantic description of relative clauses superseding the set intersection proposal quoted above; at least Chomsky's later writings rely on the set intersection analysis and on no other.)

It needs to be added that Chomsky restricts his proposal to relative constructions like *the book we ordered*. Footnote 47 of QFI gives an example of what the proposal is not supposed to cover, although there is no characterization of quite what class of relative constructions the proposal does cover: "in Navajo there are relative clauses which have, essentially, the interpretation: the dog is chasing the cat, which were fighting. These require a rather different analysis." This footnote also suggests that *the girl left and the boy arrived who*

met in Chicago and analogous examples in English which, to Chomsky, seem "at best marginal", would also, if they were non-marginal, require an analysis different from the set intersection proposal. These examples show that Chomsky is excluding cases where a single antecedent phrase links up with two or more relative phrases. Let us, with Chomsky, refrain from considering such 'split control' cases in what follows.

Consider again our example (36).

(36) je chele je puroSkar caY Sey chele Sey puroSkar pabe
"which boy which prize wants that boy that prize will-get"

As required, this example shows a biunique relation between relative phrases and antecedents – pivot control rather than split control.

Let us try to build a semantic representation for (36) using the set-theoretic apparatus used by Chomsky and on the basis of his hypothesis that the "relative clause construction […] corresponds to set intersection". In order to handle subject-object-verb relations, we define a set G (for Get) each of whose members is an ordered pair with the getter occupying the first position and the gotten occupying the second position in the pair. Thus, *boys get prizes* would under these assumptions be represented as (B, P)∈G: c[B] = c[P] ≧ 2, where B and P are the sets of boys and prizes respectively and the parenthesis-comma notation represents an ordered pair. To simplify matters, we will follow Chomsky in ignoring tense.

As a first approximation, let us define four sets in addition to G: B, the set of boys; P, the set of prizes; U, the set of things that want prizes; and W, the set of things that boys want. Under this approximation, the semantic representation of (36) is (B ∩ U, P ∩ W) ∈ G: c[B ∩ U] = c[P ∩ W] ≧ 1. This representation asserts that the ordered pair whose first member is the intersection of B (the set of boys) with U (prize-wanters) and whose second member is the intersection of P (prizes) and W (things wanted by boys) is a member of G, the set of ordered pairs whose first member gets the second member, and that the cardinality of the intersection set of B and U is equal to that of the intersection set of P and W, both being not less than 1.

A second look at this semantic representation shows that it means 'each boy who wants prizes will get some prize that boys want', whose truth-conditions differ from those of (36). We would obviously need to refine the semantic representation in order to remove this discrepancy. Unfortunately for Chomsky's set intersection theory of relative clause semantics, the adjustment which we need to make can only be made if we introduce some sort of mapping between particular boys and particular prizes and make this mapping part of our semantic representation of (36). Such a mapping, however, would involve (or be equivalent to a formalization involving) crucial use of variable binding in a way which would render the use of the set intersection device nugatory. It is impossible to give a satisfactory account of the meaning of (36) if we choose to remain within the set intersection view of relative clause semantics.

Thus, we must reject the set intersection proposal as a basis for the analysis of one-relative-to-one-antecedent relative clause constructions. To see that this does constitute a rejection of the 'spirit' and not the mere 'letter' of Chomsky's proposal, look at what Chomsky & Lasnik (1977: 447) say about the semantic vacuity of relative words in contrast to the quantifier character of question words.

It is necessary, then, to come up with a more generally usable analysis of relative clause semantics. Since, in Bangla, it is only Left Relative Clauses that offer crucial examples

choosing between Chomsky's proposal and more general solutions (given that a Right Relative never contains more than one relative phrase), we would do well to restrict our attention to left relatives so that we can take advantage of the asymmetry. If one obtains a result for double relative sentences like (36) one will be able to extend this result to sentences with only one relative; the converse, as we have just seen, does not hold.

8.8 Some other proposals for interpreting multiple relatives

One of the facts about (36) which needs to be described is its ambiguity. It can mean either 'given any pair x, y, such that boy x wants prize y, x will get y' (over a domain of boys and prizes given by the discourse) or 'for the unique pair x, y, such that boy x wants prize y, x will get y'.

One way to represent this ambiguity would be to map the sentence on to a pair of logical propositions, each representing one reading.

For the first reading we could use the logical representation (=xy: x∈B & y∈p)((x, y)∈W)((x, y)∈G), where B, P, W, G are respectively the set of boys, the set of prizes, the set of wanter-wanted ordered pairs, and the set or getter-gotten ordered pairs, and where = is a Mostowski-Tennant '(1, k)-quantifier' with k = 2. See Mostowski (1957) and Altham & Tennant (1975) for discussion of (1, k)-quantifiers; briefly, such a quantifier binds variables in an ordered k-tuple of formulae, where k is an integer. Let us call = the equational quantifier.

For the second reading, we could use the logical representation (∃!(x, y): x∈B & y∈P & (x, y)∈W)∈G, where ∃! is the unique-existential quantifier.

Another way to handle the ambiguity of (36) is to refrain from mapping (36) on to logical representations directly within one's linguistic description of relative clause semantics. One might propose, for example, that an interpretive rule subjects each relative phrase to some antecedent by marking the former as a bound anaphor, and that some combination of linguistic and extralinguistic factors makes it possible / necessary for some relative sentences to have the first reading and for others to have the second reading. This proposal would have the merit of facing the fact that not all relative structures permit both interpretations and that quite often nothing in the syntactic-lexical structure of the sentence enables us to decide what sort of logical mapping to perform. This fact shows up in the difference between *je meYe Omit raYke je bhaSaY prem nibedon korechilo Sey meYer kache Sey bhaSa pOre biSomOY hoYe uThlo* "which girl Omit Ray-OBJ which language-LOC love offer had-done that girl-GEN to that language later unbearable become-and rose" 'for girl x and language y, such that x once expressed love to Omit Ray in y, y later became unbearable to x', which can only have the second sort of reading (with a unique-existential quantifier), and *je ja paY Se ta caY na* "who what gets (s)he that wants not" 'for person x and thing y such that x gets y, x doesn't want y', which can only have the first sort of reading. The reason why the first sentence must have the unique-existential quantifier reading is that every literate speaker of Bangla knows that Omit is a character in Tagore's famous novel *SeSer kobita* 'the last poem', and that the girl in question is Ketoki. The reason why the second sentence must have the (1, k)-quantifier reading is that the sentence is too much of a cliché to be able to refer to a particular person.

There are independent reasons why the task of disambiguating sentences with respect to reference and quantification should be placed at the interface between linguistic description and the study of other faculties of the mind. Investigations of quantification in natural language

have abundantly shown that quantification is subject to the indeterminacies and idiosyncrasies characteristic of lexical items. For a perspicuous statement of this finding, see McCawley (1977). Robert Fiengo (1978) argues that, contrary to hypotheses current in much lexicalist work, the interpretation of reciprocals must be relegated to 'second'-phase semantic interpretation, which works at the interface linking the lexicon to common-sensical and other sources of real-world knowledge. Consider also the way in which expressions of universal quantification are embedded in such words as *entire, complete, per* ('for each'), *apiece*, and so forth; surely it cannot be the task of the interpretive rules of sentence grammar to unravel the irregular workings of such words.

On these grounds, I hypothesize that Bangla sentence grammar does not map a structure like (36) directly on to an unambiguous representation of its meanings (such as mathematical representations of the sort discussed above), but rather specifies the relation of bound anaphora which binds each relative phrase to some antecedent phrase. It is convenient to couch this hypothesis in terms which permit direct comparison with a 'logicalist' alternative analysis. Thus, I postulate an interpretive rule in Bangla grammar which maps each relative clause on to a partly logical and partly linguistic representation in which the relative element J has been consistently replaced by a logical variable element VBL, and another interpretive rule which maps each antecedent-containing clause on to a representation in which the antecedent element, whether the T determiner element of T-Words (Sequent Words) or the word *Emon* 'such', has been consistently replaced by a logical archi-operator element written BIND. It is this archi-operator BIND which will, in later stages of the 'computation', get spelled out as the appropriate operator – the (1, k)-quantifier '=' or the unique-existential quantifier '∃!' – and linked up with the variable in mathematically appropriate form; but that phase of the computation is outside the scope of Bangla linguistics and perhaps of linguistics as such.

The upshot is that I am proposing (60) as the first-phase semantic interpretation for the boy and prize sentence (36).

(60) VBL chele VBL puroSkar caY BIND chele BIND puroSkar pabe

One problem with this proposal is that it offers no means of pairing up appropriate relatives with appropriate antecedents. I return to the problem of pairing in section 8.11, where I argue, in effect, that the problem is insoluble within the grammar.

Let us call this hypothesis the Grammatical Interpretation Hypothesis (GIH) and the alternative sketched earlier the Logical Interpretation Hypothesis (LIH). Despite the difficulty of deciding between competing semantic analyses on empirical grounds, I will now try to show some advantages of GIH over LIH. This comparison of the theories will be unaffected by the fact that GIH cannot pair up the right relative phrase with the right antecedent. LIH has no algorithm for that task either. For reasons of space, I will not carry out a full-fledged theory comparison, but will simply offer an exposition of GIH in full detail. To the extent that GIH, which has less 'power' than LIH, can do a considerable amount of work in the semantics of Bangla, the extra 'power' of LIH will be seen to be unnecessary.

At the level of particular-language analysis, the following section presents a GIH analysis of Bangla indefinite forms, with the suggestion that LIH fits the data less well. On this purely descriptive plane one might as well say that, for present purposes, 'GIH' *is* the view that the sentence grammar of Bangla describes the semantics of relative clause constructions without distinguishing 'equational' from 'unique-existential' readings, while 'LIH' is the view

that Bangla sentence grammar in its semantic component distinguishes 'equational' from 'unique-existential' readings of relative clause constructions.

At the same time, at the level of general linguistics, I am using the terms 'GIH' and 'LIH' to designate two different directions that linguistic theory-building might take. I am trying, unemphatically, to suggest that GIH is the more suitable of the two. The LIH direction, as I see it, is one in which linguistic (sentence-grammar-defined) semantic representations are set up to be as far as possible 'camera-ready' for a one-to-one mapping from these representations on to unambiguous logical formulae. The GIH direction – again, as I see it – is one in which linguistic semantic representations are derived by rules sensitive only to syntactic information and to information from inflectional morphology and the 'second lexicon' (entries for grammatical morphemes), so that the mapping from these linguistic semantic representations on to unambiguous representations of meaning is relatively complex.

It is important to see that it is a personal decision on my part to interpret the result of the following section (the manifest superiority of the GIH over the LIH approach to Bangla indefinite forms) in the light of my particular demarcation between GIH and LIH as general approaches. This subjective decision, with which some linguists will, no doubt, disagree, may be refuted by later research in cases (if any) where empirical considerations favor what I am calling LIH rather than what I am calling GIH. In the light of such cases, if and when they emerge, it will be necessary to draw a different line between properly linguistic semantics and 'semantics beyond sentence grammar'.

8.9 The grammatical interpretation of indefinite forms

In order to distinguish GIH from LIH and make it easier to choose between them, I shall reformulate GIH in a way which its logic does allow but the logic of LIH does not – as a lexical rather than a derivational hypothesis.

Since LIH is committed to the project of mapping syntactic structures on to unambiguous logical forms within grammatical derivations, it must use rules that scan phrase-markers, find triggering elements, and produce new phrase-markers. Since GIH embodies the view that the grammar is responsible only for whatever meaning appears in function words and structures, GIH can quite naturally allocate the job of marking relative words as VBL and antecedent words as BIND to the system of lexical features in the 'second lexicon' (the set of function words and affixes). Thus, one may restate GIH so that no rule needs to operate on (36) to yield (60) – by ensuring that the structure of (36) itself contains the semantic feature specifications +VBL and +BINDER:

(61) je chele je puroSkar caY Sey chele Sey puroSkar pabe
+VBL +VBL +BDR +BDR

The features are already present in the entries for *je* and *Sey* and enter the phrase-marker when these elements are inserted, on this view. LIH, which is committed to unambiguously representing all predication and binding relations in its output, cannot transfer its work to the lexicon, since *je* and *Sey* are ambiguous in crucial respects.

Thus lexicalized, GIH becomes a hypothesis about the structure of lexical entries for Phoric Words of at least three sorts: J-Words (relative words), T-Words (sequent words), and

the E-Word *Emon* 'such'. Given the kinship between J-Words and K-Words shown on other grounds, one expects K-Words also to have the feature +VBL. One further expects 'K-o-Words', indefinite words, to contain this feature specification in their interrogative (K-)Word half. Let us postulate that these expectations are met – let us, in other words, assign the feature +VBL to K-Words, and to the K-Word half of K-o-Words. We now no longer need to use OPEN as an independent feature. The words we were marking as +OPEN now become simply those +VBL words which do not have an indefinitizing second half. We still do need OPEN as a derivative feature, however, in order to be able to state such generalizations about Open Words as their inability to take the Emphasizer /o/. So, we can postulate a lexical redundancy rule marking every [+VBL]-throughout word as [+OPEN] and all other words as [–OPEN] (unless we prefer to adopt a general convention to perform the latter task, as in one of the options inconclusively considered in Chomsky's *Aspects*).

Let us look at indefinite words a bit more closely. The first half of a K-o-Word, the K-Word half, has a semantic feature specification now which says that it is, in effect, a variable. The semantic difference between an interrogative (K-)Word and an indefinite (K-o-)Word must therefore lie in the second half of the word, in the indefinitizing /o/ element.

Judging from sentences like *ram kothaW giyeche* "Ram somewhere has-gone" 'Ram has gone somewhere' and *ram kothaW jaY ni* "Ram anywhere goes not-PERF" 'Ram hasn't gone anywhere', the contribution of this /o/ element to the meaning of the sentence containing it seems to be something in the nature of an existential quantifier. For example, one could informally represent the meaning of the above sentences as *for some x: x a place, Ram has gone to x* and *not (for some x: x a place, Ram has gone to x),* respectively.

Accordingly, the semantic part of the representation of the indefinitizing /o/ – whatever the form of this representation may be, an entry in some mini-lexicon, a rule for forming indefinite words from interrogatives, or something else – should say that this element is an existential quantifier binding the variable associated with the K-Word to which the /o/ is suffixed.

Despite the plausibility of this analysis, I must surround it with caveats. It may be true that the indefinitizing /o/ does nothing but function as an existential quantifier of sorts. But certainly other elements in the position of right adjunct to a K-Word also function, at least in part, similarly. /SOb/, which elsewhere simply pluralizes or means 'all', is such an element. Only a limited subset of K-Words take /SOb/ as a suffix. Expressions like *ki-SOb* "what-all", *kara-SOb* "who-PL-all", etc. have readings which verge on the exclamatory but primarily assert indefiniteness: *jiten fon korechilo, ki-SOb Ongko koSe dite hObe mone koriye diyeche* "Jiten phone did, what-all problems calculate-and to-give will-be-necessary remind-and has-given" 'Jiten called, he reminds you that you're supposed to do some maths problems', the implication being that the speaker doesn't know and perhaps doesn't care what the problems are but presumes the addressee knows what Jiten is talking about. Other examples, like *ki-SOb baje kOtha!* "what-all nonsensical talk!" with typically an HH L L L L intonation contour, are more obviously exclamatory. It is not clear that one should characterize *SOb* used in this fashion as having a hidden existential side to it, especially in view of the ambiguity noted. For even K-Words sans adjunct, in highly restricted, poorly explored environments, take an 'indefinite' reading: *kothakar kon bOd cheler SOngge miSe khisti-kheuR Sikheche* "what-place-GEN which rotten boy-GEN with mix-and swear-words has-picked-up" 'He's been with some rotten kid from somewhere and picked up swear-words'. The only difference between such 'environment-induced' or /SOb/-induced indefiniteness and the indefiniteness conveyed by /o/ is that the

latter, but not the former, can do without back-up from neutral intonation and appropriate context, and doesn't constantly totter on the brink of exclamationhood.

I hypothesize that the K-Word with its variable character opens a semantic gap which immediately gets filled by a neighbouring /o/, if present, or by other, less reliable, adjoining or suprasegmental gap-fillers, perhaps with support from the given or assumed context. This hypothesis plays up the variable and permits more latitude in finding or imagining an operator that shall bind this variable.

Such a semantic analysis of indefinite words seems rather natural. In languages as different from one another as Hindi, Sanskrit, Japanese, German, French, English, and Bangla, we find some, or most, indefinite words to be morphologically made up of an interrogative word plus some increment. Given analogous situations in derivational morphology – for example, the relation between *ferocious* and *ferocity*, between *electric* and *electricity* – the unmarked semantic description (the null hypothesis) is one which assigns a composite interpretation to the composite morph sequence and preserves as much correspondence as possible between the morphemic composition and the semantic composition of the composite form. Surely no cogent argument has been offered for not applying this method to the case of indefinite expressions as I have done in the above proposals.

The Logical Interpretation Hypothesis, in contrast, would require some device or devices for mapping sentences which contain indefinite expressions on to logical formulae containing existential quantifications written with the quantification outside (conventionally to the left of) the propositional function which contains one or more variables bound by the quantification. It seems clear that LIH would be quite capable of handling cases with indefinite words, which occasion no ambiguity. But cases involving /SOb/ and/or intonation plus non-linguistic context would be beyond the grasp of the principles constituting LIH. The best that LIH could do would be to construct one-many mappings on to a variety of logical or para-logical formulae, some with existentials and others with ad hoc operators denoting exclamation, vagueness, or other elements involved in readings which depart from existential interpretations. GIH, with its emphasis on lexical peculiarities combining with other factors to make up these meanings, is better at handling the subtle cases.

Here ends the explicit comparison between GIH and LIH; I regret that I have not given a proper formulation of the LIH analysis which I have been unfavourably comparing with the GIH analysis. The following analyses within the GIH framework, to the extent that they are adequate, imply an argument against the more cumbersome analyses of the phenomena under LIH assumptions; however, I won't spell out this argument or the cumbersome LIH analyses, mostly for reasons of space. See May (1977) for an extensive example of LIH-inspired work; within that framework May's work is quite careful and expresses many insights of general interest.

8.10 Constituent Questions

One assumption about constituent question semantics embedded in the above account of indefinite forms is that, since interrogative words are specified + for the feature VBL, a clause containing such words counts as a linguistic equivalent to what logicians call a propositional function or open sentence – a formula which leaves one or more variables 'free', unbound by any operator.

Earlier and worthier investigators have made the same assumption. Logician Felix S. Cohen (1929: 359) proposes interpreting a question like *What is mortal?* as the propositional function *x is mortal*. Philologist Otto Jespersen (1965: 303) says that in constituent questions 'we have an unknown "quantity" exactly as in an algebraic equation; we may therefore use the well-known symbol *x* and the term *x-question* for a question aiming at finding out what *x* stands for'. I have found no attempts in the literature (let alone successful ones) to cast doubt on the legitimacy or usefulness of this assumption, on which I will therefore base a theory of the meaning of constituent questions.

A direct question (in particular, a direct constituent question) in most situations invites an answer, in some sense of 'invite'; so much so that Katz & Postal (1964), Jespersen (1965), and other authors have strengthened the invitation idea into the idea that *I request that you answer* is a general meaning-frame for the use of questions. Baker (1970) has pointed out that this stronger idea has no application to indirect questions; cf. the absurdity of *I wonder I REQUEST THAT YOU ANSWER why I'm glad*. One may also doubt its usefulness for direct questions, despite its initial appeal in this case. In all sorts of behaviour (even non-human behaviour – e.g. a bird too young to fly 'inviting' its mother to feed it, opening its hungry mouth), organisms leave a gap, overtly, for an interaction-partner to fill. Must there be an 'act of invitation to fill the gap' distinct from the 'act of leaving a gap' and governing the latter?

Turning to issues within the linguist's special province, we note that, while direct questions do leave an overt gap, indirect questions do not.

(62) chobiTa kar tola?
"picture-item who-GEN? taken"
'Who took the picture?'

(63) chobiTa kar tola bole ram oder janiyeche?
"picture-item who-GEN? taken as Ram them has-told"
'Who did Ram tell them took the picture?'

(64) ram oder janiyeche chobiTa kar tola
"Ram them has-told picture-item who-GEN? taken"
'Ram told them who took the picture'

Sentences (62) or (63), which are direct questions, 'invite' an answer; but (64), which contains an indirect question, does not invite an answer. An adequate theory of the meaning of constituent questions must describe this difference between Indirect and Direct.

As a step towards such a theory, we note that it is the entire clause which, in some sense that needs to be made precise, 'chooses' to be a direct question or to be an indirect question, although one might a priori imagine that it would be the individual questioned NPs which would serve as the loci of interrogation in a constituent question.

Thus, in (65), both *kon chobiTa* and *kar* take part in the indirectness of the indirect question *kon chobiTa kar tola*. And in (66) both *kon chobiTa* and *kar* take part in the directness of *kon chobiTa kar tola?*, a direct question. Never does one NP have a direct question effect while its clause-mate NP goes indirect. (It is not hard to construct examples with that property; those examples are always ungrammatical.)

(65) ram oder janiyeche kon chobiTa kar tola
"Ram them has-told which? picture-item who-GEN? taken"
'Ram told them who took which picture'

(66) kon chobiTa kar tola bole ram oder janiyeche?
"which? picture-item who-GEN? taken as Ram them has-told"
'Who did Ram tell them took which picture?'

To use a slightly different example in order to make the same point, consider the sentence *ke kake SaMtar kaTte dekhlo*? "who? whom? swim to-cut saw" 'Who saw whom swim?', which is a direct question and in which *ke* "who?" and *kake* "whom?" both have a 'direct question' effect, and the sentence *ke dekhlo ke SaMtar kaTche*? "who? saw who? swim cutting-is" 'Who saw who was swinming?', which is a direct question containing an indirect question and in which *ke* "who?", which has a 'direct question effect', is no longer a clause-mate of, but is rather an aunt of, the NP *kake* "whom?", which has an 'indirect question effect'. The reason for the difference between these sentences, in our terms, is that in the former sentence *kake* and *ke* are in the same clause, while in the latter sentence only the second *ke* (not the first *ke*) occurs in the embedded clause.

These simple facts, incidentally, sharply distinguish the semantics of interrogation from that of (universal or existential) quantification. Notice that theories which posit a 'question quantifier' to handle the interpretation of questions, like the theories of Chomsky (1978) or Reichenbach (1947) discussed in chapter 10, must be rejected unless they are augmented by adding some explanation for the above-noted special features of questions – features not shared by sentences with universal or existential quantification.

The generalization which emerges is that the Indirect or Direct status of a constituent question gets decided at the level of the clause containing the question words, and not at the level of particular questioned NPs. In order to account for this generalization, and also in fulfillment of the need (noted in chapter 7 where we found the Complementizer Transfer Hypothesis and the No Complementizer Hypothesis to be observationally inadequate and where we therefore had to accept the Semantic Matching Hypothesis) for a semantic rule which relates the Interrogative Words in an S′ to the COMP of that S′, let us postulate the following rule of construal.

(67) Interrogative Construal
S.D.: an S′ which dominates, without any lower S′ also dominating, zero or more +INT determiners and a +INT complementizer
S.C.: construe all the determiners with the complementizer

I assume that 'to construe' is 'to build a construal for', a 'construal' being a supra-phrase-structural formal object associated with a syntactic phrase structure, much as a suprasegmental contour is associated with a phonological segmental structure. A construal built by (67) is an ordered pair. One member of this ordered pair is a variable – x, the independent variable – associated with the complementizer. The other member is an ordered n-tuple – (y_1, y_2, .., y_n), the dependent variables – associated with the determiners. The ordering, both in the pair and in the n-tuple, is not intrinsic to the construal as such, but mirrors the fact that each item in the construal is associated with a syntactic segment. Since the syntactic segments are linearly

ordered, so are their images in the construal.

The analogy between the ways in which a superimposed grammatical process and a suprasegmental contour are related to their respective segmental substrata is due to John Costello (class lecture, 1975).

(67)'s functioning is governed by the Condition on Analyzability, which has the effect of making a rule that mentions a Specifier apply to the minimal +N phrase dominating that Specifier and not immediately dominated by another +N phrase. This formulation, which is due to Sag (1976), Woisetschlaeger (1976), and May (1977), does not quite meet the needs of (67), for in (67) it is not only each DET which – as provided for by the condition – 'carries its NP or AP with it', but also the COMP which – as Specifier of S′ (May 1977) – 'carries its clause with it', although S′ is not a +N node, and although in many cases S′ is immediately dominated by S. The remedy is not to substitute some other set of categorial features for +N in the formulation of the Condition on Analyzability, since, for one thing, it is likely that there is no natural class comprising all and only the category types N, A, S and specifiable by a categorial feature bundle, and, for another, the immediate dominance proviso does not work for S′, which is often immediately dominated by S which shares its categorial feature values. If the needs of (67) are any indication, the solution lies in the direction of defining a notion 'hypophoric' such that all and only hypophoric category types can have phoric specifiers. This notion, as desired, brings the category types N, A, and S together in opposition to P and V. We now tentatively restate the Condition on Analyzability thus: if a rule mentions a specifier, then the rule applies to the lowest hypophoric node which dominates that specifier but is not immediately dominated by a +N node.

It is interesting that the spirit of the condition – the idea that the (cross-)referential status of a phrase is borne by its specifier – accords with the findings of ongoing psycholinguistic research by Roger Wales. Wales (personal communication) finds that the fundamental task that a child faces in acquiring syntax is that of framing lexical items in a format constituted by deictic specifier elements.

The present formalization of construal differs from Kuno & Robinson's analysis (1972) of constituent questions as follows. In their analysis, in effect, the interrogative COMP is coindexed with the interrogative phrases in the clause. In the present analysis, the COMP is not coindexed but merely construed with the determiners of the interrogative phrase. At least two considerations favour the present analysis rather than Kuno & Robinson's account.

First, wherever A is grammatically coindexed with B, current proposals lead one to expect that either the extension or at least the intension of A will be necessarily identical to that of B. We see necessarily identical extension in *John$_i$ lost his$_i$ way* and necessarily identical intension in *The blue panther$_i$ hates the pink one$_i$*. However, in (67), one must preserve the distinction between the semantic gaps denoted by any y_i and any y_j distinct from it. Thus, it will not do to coindex x with all the dependent variables.

Second, theoretical elaboration to be provided in section 8.11 will use the independent variable in the construal built by (67) to mediate between the dependent variables and their antecedent(s) outside the S′. For this mediated binding to take place efficiently, x itself needs to be a non-binder. Under current conventions, coindexing means binding. To use the felicitous wording proposed by Ann Reed (1975), binding entails 'congruity of reference claims', of which coreference is a special case.

These objections clearly have no force against Baker's (1970) solution in terms of multiply coindexing the interrogative complementizer (to reformulate his account with our

current descriptions in mind) with each of the interrogative phrases to yield outputs of the form …$COMP_{ijk}$…$phrase_i$ …$phrase_j$ …$phrase_k$…, thus avoiding the undesirable lumping together of the dependent variable positions. But Baker provides no theory of indexing which would enable multiple indices like these to hold their own against or alongside the standard 'congruity of reference claims' use of indices. More importantly, Baker's notation and the notions underwriting it do not provide for the mediated binding of section 8.11. I therefore find it appropriate to use a formalization markedly different from Baker's, in order to stress the difference rather than the similarity between my proposal and his.

Returning to the difference between direct and indirect questions, we now pose this problem in terms of how the direct/indirect status of a +INT COMP is determined. For, given a criterion suitable for this purpose, we may extend it to the Determiners construed with this COMP, and thus to the phrase nodes immediately dominating DET.

The most simple-minded solution to the problem thus posed would be to say that an S′ is a direct question if it has a +INT COMP and is a root S′ and that an S′ is an indirect question if it has a +INT COMP and is a non-root (an embedded) S′. This null hypothesis adequately handles certain interesting facts.

In particular, it accounts for the fact that the structure *chobiTa kar tola* "picture-CLA who-GEN? taken" 'who took the picture(?)' counts as a direct question in *chobiTa kar tola bole ram oder janiyeche*? 'Who did Ram tell them took the picture?' but as an indirect question in *ram oder janiyeche chobiTa kar tola* 'Ram told them who took the picture'. In the first sentence, the sequence *chobiTa kar tola*, being the S object of the P *bole*, does not itself have a COMP, so that its interrogative NP *kar* must place its +INT DET under the jurisdiction of the first +INT COMP overhead, i.e. the +INT COMP of the entire S′ *chobiTa kar tola bole ram oder janiyeche*. Since this S′ is a root S′, it follows that *kar* "who-GEN?" must serve as a direct question, via the COMP of the root S′. In the second sentence, the sequence *chobiTa kar tola*, since it is an embedded S′, has its own COMP, which must be +INT to match the +INT DET in *kar*; since the S′ *chobiTa kar tola* is an embedded S′, it must function as an indirect question.

However, the hypothesis that a question is indirect iff embedded is incompatible with the plausible claim (which I, for one, would make) that, in *ram oder janiyeche chobiTa kar tola* 'Ram has told them who took the picture', the S′ *chobiTa kar tola* is coordinated to the S′ *ram oder janiyeche* and is therefore simultaneously a root sentence and an indirect question. We need an alternative account of direct and indirect questions which makes no essential reference to the notion 'embedded'.

I propose that a question clause is indirect if and only if it functions as a Thematic Relatum (as an 'Argument', in the parlance of classical predicate calculus) to a Predicate.[15] Thus, in the example discussed above, *chobiTa kar tola* is the Theme of *janiyeche* although it is not a syntactic Object or Complement.

This proposal, unlike the indirect-iff-embedded hypothesis, explains why indirect questions leave no overt gap. That a question clause is syntactically embedded entails nothing directly about its gap-leaving behaviour. But if we know that a question clause functions as a Relatum to a Predicate, we expect that the question with its gap will be assimilated into the

[15] A 2020 note: Contemporary readers need to be reminded that this dissertation was written before 'theta-roles' and 'theta-marking' became standard fare. I was working in Australia, and had only **heard** of the Pisa lectures: I had **access** to them, and to the GB literature, only after submitting this thesis.

predicational structure so that the gap will cease to be an overt one.

In some sentences, the indirect question is a Relatum only by proxy – for example, *chobiTa kar tola ram oder ta janiyeche* "picture-item who-GEN? taken Ram them that has-told" 'Ram told them who took the picture', where the NP *ta* "that" is the Theme of *janiyeche*. In this sentence it seems to me reasonable to say that the S′ *chobiTa kar tola* 'specifies' the meaning of the NP *ta* in a sense of 'specify' related to the sense explained by Joan Bachenko in her work on English pseudo-cleft structures (Bachenko (1976), Bachenko & Teller (forthcoming)). As in Bachenko's work, this S′ inherits the functional relations which the NP it specifies is party to. Since the NP is the Theme of *janiyeche*, the S′ specifying the NP also 'is the Theme of' *janiyeche*.

Both of the hypotheses I have considered – the indirect-iff-embedded hypothesis and indirect-iff-Relatum hypothesis – satisfactorily account for the way in which the 'question effect' of questions 'travels up' a chain of /bole/'s, as in *ke rag koreche bole ramer dharona bole SEmer biSSaS*? "who? anger has-done as Ram-GEN understanding as Shyam-GEN belief" 'Who does Shyam believe Ram thinks is angry?'. A clause governed by P is an S and not an S′; therefore, a K-Word in such a clause must be construed with the lowest available COMP, which, in the sentence given, is the COMP associated with the entire S *ke rag koreche bole ramer dharona bole SEmer biSSaS*. Since both the indirect-iff-embedded proposal and the indirect-iff-Relatum proposal refer to S′ and not to S, they both adequately describe the behaviour of P-governed S. A P-governed S, not being an S′, cannot by itself be a question (direct or indirect) even if it contains one or more K-Words: only the smallest S′ of which this S forms part may count as a direct or indirect question.

Perhaps this point will become clearer if I give an example where an S′ containing an S + *bole* structure is an indirect question: *ke rag koreche bole ramer dharona ami ta jani na* "who? anger has-done as Ram-GEN understanding I that know not" 'I don't know who Ram thinks is angry'. Here the question S′, *ke rag koreche bole ramer dharona*, specifies *ta*, and inherits the role of Theme of *jani na*, which makes it an indirect question.

The indirect-iff-embedded hypothesis and the indirect-iff-Relatum hypothesis are related in the following fashion. If a clause is embedded **as a complement**[16], it will end up being a Relatum, since that is how predication works. But if a clause is a Relatum, it is not necessarily embedded.

The theory of constituent question semantics developed in this section has served to clarify somewhat the previous section's analysis of indefinite forms. The whole picture will become even clearer in the following section, which takes up the semantics of relative clauses.

8.11 Relative clauses

In this section I will first extend the Semantic Matching Hypothesis to cover relative clauses as well as constituent questions. Then I will show that this analysis of the semantics of relative clauses makes it possible to account for phenomena which cannot be handled within the

[16]The words in boldface are a 2020 addition. Had I noticed this twist in 1980, I would have realized (a) that the case of an interrogative clause embedded as an **adjunct** needs separate analysis – and (b) that an interrogative constituent in such a clause does take wide scope. A crucial omission on my part.

alternative grammar for relative clauses proposed by Chattopadhyay (1976a, b).

If one takes the rule of Interrogative Construal (67) and substitutes α for + everywhere, one obtains a rule of Open Specifier Construal which applies to constituent questions when α is taken to be +, and to relative clauses when α is taken to be –.

(68) Open Specifier Construal

S.D.: an S′ which dominates, without any lower S′ also dominating, zero or more αINT determiners and an αINT complementizer
S.C.: construe all the Determiners with the Complementizer

One carries over the formal definition of construing which was used for (67). But there is now the extra task of handling the antecedent-anaphor relation between, say, the antecedent phrases *Sey chele* and *Sey puroSkar* and the anaphoric phrases *je chele* and *je puroSkar* in *je chele je puroSkar caY Sey chele Sey puroSkar pabe* "which boy which prize wants that boy that prize will-get". At first sight it seems that we simply need to add another rule of interpretation, (69), which subjects anaphors to binding by antecedents.

(69) Relative Binding

S.D.: an S′ with which a construal $((y_1, y_2, \ldots, y_n), x)$ is associated, embedded in a larger S′, the variables all being –INT
S.C.: bind the dependent variables, one antecedent per variable

The structural change is unspecific because, under the Grammatical Interpretation Hypothesis, sentence grammar cannot pick out the right controllers.

However, (69) fails to account for several facts which a grammar of relatives must cover. Since J. Chattopadhyay's grammar of relatives proffers only a drastically reduced form of (69) (her rule, as shown in section 8.5, accommodates only cases with one relative phrase in the relative clause, which lines up with one sequent phrase in the matrix), the following arguments against (69) hold a fortiori against her analysis.

The first fact which poses a problem for (69) is split control, which appears in sentences such as *e deSe jOto chatro jOto boy pORe o deSe tar ceYe beSi-i dekhbo bole aSa kori* "this country-LOC as-many student as-many book read that country-LOC that-GEN than more-EMP will-see as hope do" 'I rather hope that I will see more students reading more books in that country than in this country' (the translation does the comparison 'backwards', as English requires). In this sentence, which I will henceforth call the student and book sentence, the NP *tar*, a single sequent phrase, cannot be matched with either of the relative phrases *jOto chatro* 'as many students' and *jOto boy* 'as many books'; nor does it seem reasonable to match it up directly with the ordered pair (*jOto chatro*, *jOto boy*) by means of a special anaphora rule which constructs the ordered pair for this purpose, since to take that route is to forget that one is constructing precisely the same pair as one already had to construct for the purposes of the construal rule (68), which had nothing to do with binding by the sequent antecedent.

If we call the phenomenon in the boy and prize sentence Pivot control and the phenomenon in the student and book sentence Split control, then we must speak of Pivot plus Split control in the following sentence: *ager juge* je_h *rOkom biSSobiddalOYe* je_i *rOkom*

oddhapok je$_j$ rOkom chatroke je$_k$ rOkom biSOY pORaten ajkal Se$_h$ rOkom biSSobiddalOYe Sei rOkom oddhapok ar ta$_{jk}$ pORan na "former period-LOC which$_h$ type university-LOC which$_i$ type professor which$_j$ type student-OBJ which$_k$ type subject taught nowadays that$_h$ type university-LOC that$_i$ type professor any-more that$_{jk}$ teach not", roughly 'The sort of professor in the sort of university that used to teach x type students y type subjects no longer does so'. In this sentence, which I will henceforth call the nightmare sentence, the antecedent phrases *Se rOkom biSSobiddalOYe* "that type university-LOC" and *Se rOkom oddhapok* "that type professor" pivot-control their anaphors *je rOkom biSSobiddalOYe* "which type university-LOC" and *je rOkom oddhapok* "which type professor" respectively, while the antecedent phrase *ta* split-controls the remaining relative phrases *je rOkom chatroke* "which type student-OBJ" and *je rOkom biSOY* "which type subject". The existence of such Pivot plus Split Control is the second problematic fact for (69).

The third such fact is the unity of split — the fact that you never get a sentence with one antecedent phrase split-controlling a couple of relative phrases and another antecedent phrase split-controlling another couple of relative phrases. A semantically plausible example of such a sentence would be **o deSe jOto biSSobiddalOYe jOto chatro je rOkom oddhapoker kache je rOkom biSOY poRte paY e deSe tOto ta paY na* "that country-LOC as-many university-LOC as-many student which type professor-GEN with which type subject to-study get this country-LOC that-many that get not", supposed to mean 'It is not the case that as many students at as many universities in this country get to study the same sort of subjects with the same sort of professors as is the case in that country'. Of course, the ungrammaticality of this sentence is predicted by (69) as it stands, but the point is that if we make a simple-minded attempt to remedy the obvious inability of (69) to cover Split control we are likely to end up with a rule which predicts that a sentence can have two instances of split control side by side – a false prediction, as our doubly split example shows. The doubly split example also shows that it would be wrong to abandon the problem altogether and claim that it is not a problem on which grammatical principles are to be brought to bear. The fact that only one Split is allowed per sentence surely pertains to sentence grammar, even if sentence grammar cannot tell precisely which antecedent is to bind which anaphor(s).

Having noted these observational problems with (69), we further note, on the plane of descriptive adequacy, that (69) is insensitive to the complexity of the output of the construal rule of Open Specifier Construal. (69) merely pairs up antecedents with relative phrases, as if Open Specifier Construal had not operated to set up a formal object which places the relative phrases under the jurisdiction of the COMP of their S′. It is merely accidental, given the formulation of (69), that the relative phrases are inside a particular S′ while their antecedents are outside this S′. As it stands, (69) in fact even permits us to have *take* "him/ her" bind *je* "who" in *je take cene* "who him/her knows", a horrifying degree of permissiveness.

Taking all these problems into account, let us replace (69) by (70).

(70) Relative Binding

S.D.: an S′, whose COMP is –INT and with which a construal $((y_1, y_2, \ldots, y_n), x)$ is associated, embedded in a larger S′

S.C.: make it so that each variable is bound by at least one antecedent in the matrix S′ and any single antecedent in the matrix directly binds at most one variable.

Obviously this rule by itself will not suffice. We must spell out the conventions of indirect binding which govern its application.

If every dependent variable is bound, a reasonable convention would deem x itself to be 'indirectly' bound, although not by an antecedent in particular. Let us call this the Overbinding convention; x will be said to be overbound if it is bound under this convention.

Another convention shall dictate that, if an antecedent binds x, then that antecedent counts as indirectly binding (in this case, 'subbinding') every dependent variable which is not otherwise bound. I will call this the Subbinding convention and variables bound under it subbound variables.

The point of these conventions should be clear. Between them, they ensure that no variable is left unbound if there is even a single antecedent (potential antecedent) in the matrix. I will return to the question of whether this result is quite what we want. But first I must show how (70) and the two conventions solve the problems noted before.

Split control is now easy. In the student and book sentence, the phrase *tar* "that-GEN" now directly binds x, the variable associated with the COMP of the embedded S′ *e deSe jOto chatro jOto boy pORe* "this country-LOC as-many student as-many book read", which has acquired this x through rule (68). By directly binding x, *tar* gets to subbind *jOto chatro* as well as *jOto boy*. As this example shows, subbinding has the interesting property that the subbinding item manages to be 'equivalent' to all (in this case, to both) the subbound items without triggering application of the axiom that, if B and C are both equivalent to A, then B is equivalent to C. This shows that subbinding is not, technically, an equivalence relation. No wonder, since it is irreflexive and asymmetric to start with!

Even Pivot plus Split control is now rendered manageable. In the nightmare sentence cited above with the indices h-i-j-k, we will now say that antecedent h and antecedent i bind relative phrase h and relative phrase i respectively (or, more rigorously, bind the y-variables associated with these relative phrases) and that antecedent jk binds x and thus gets to subbind the y-variables associated with the otherwise unbound relative phrases j and k.

Finally, the unity of split now follows from the fact that at most one sequent phrase can bind x and the fact that exercising split control is tantamount to binding x.

Thus, an analysis of relative structures incorporating Relative Binding and the conventions of Overbinding and Subbinding accounts for the seemingly complex facts which a relative-sequent pairing analysis (such as the preliminary version (69) of Relative Binding, or the SNP analysis) cannot handle.

Let us return now to the question of the desirability of the result which we now have — the result that no variable is left unbound if there is even a single potential binder in the matrix. Surely, one may argue, there are cases where we want to block binding, as in *je biSSobiddalOY birotto dekhiyechilo Sey chagol Ekhon dhuliSat* "which university heroism displayed that goat now dust-reduced" 'The university-goat which displayed heroism is now reduced to dust', where the speaker hasn't made up his or her mind about the nature (goat? university?) of the pivot noun which is shared between the relative phrase and its pivot-controller. On the basis of such examples, one is tempted to build safeguards. The problem is that safeguards won't be foolproof. One can say things like *je chatro Sohid hoYechilo Sey salamke kew mone rakhe ni* "which student martyr became that Salam-OBJ anyone mind-LOC keep not-PERF" 'No one remembers that Salam which student became a martyr' – bad English but good Bangla. Surely one does not wish to construct a purely grammatical algorithm capable of recognizing the possibility of equating *salam* with *chatro* and the impossibility of equating a university with a

goat.

Perhaps slightly more significant for the grammar would be the problem which one might pose on the basis of examples like *ram jake ja Sikhiyechilo Ekhon Se ta bhulte caY* "Ram whom what taught now (s)he that to-forget wants". This sentence can mean either 'The person$_i$ Ram taught it$_j$ to now wants to forget what$_j$ Ram taught that person$_i$' or 'Ram now wants to forget about having taught what$_j$ he did to the person$_i$ who$_i$ he taught it$_j$ to'. On the first reading, *Se* 'that person' binds the y-variable of *jake* 'whom' and *ta* 'that' binds the y-variable of *ja* 'what', so that the x-variable is deemed overbound. On the second reading, *Se* 'that person' is used as a free anaphor to *ram* and binds nothing, while *ta* 'it' binds the x-variable and thus subbinds both of the y-variables. On the basis of this example one might object to my analysis, saying, 'The analysis fails to ensure the derivation of the second reading alongside that of the first reading.'

Such an objection would rest on a misunderstanding which perhaps needs to be forestalled. The mere fact that a relative clause has *n* relative phrases and its matrix clause *n* potential binders does not lead (70) to employ all the potential binders, each to bind one relative phrase. Phrases in the matrix clause which qualify as antecedents also have other options (including free anaphorhood), which they may exercise instead of serving as binders. All that is needed for a well-formed derivation is that at least one potential antecedent 'consent' to work as a binder; it will then (in the case of several relative phrases waiting to be bound) split-control all the y-variables.

How to ensure well-formedness is a different question. One way would be to make rule (70) obligatory, so that – if a relative clause finds no antecedents – the rule, having failed to apply to an eligible structure, will star the structure, as obligatory rules in general do. However, this device would not catch those relative clauses which get saddled with inappropriate binding, as in, say, **rOmeS jake cinto ta jani na* "Romesh whom knew that know not" 'I/we don't know the thing whom Romesh knew', where the human *jake* is bound by the non-human *ta*, but (70) is satisfied. Since some sort of checking, possibly involving joint action of linguistic and pragmatic factors, will be necessary in such cases, perhaps we can rely on this checking process to guarantee that the application of (70) will be obligatory in net effect without stipulation of such obligatoriness. I leave the question open, in view of this possibility. (Some readers may not see the problem being addressed here, and may need to be reminded that substituting interrogative *kake* "whom?" for the relative *jake* "whom" would turn the initial three-word sequence into an indirect question and would thereby remove the ill-formedness of the sentence under scrutiny. In other words, one **can** say *rOmeS kake cinto ta jani na* "Romesh whom? knew that know not" 'I/we don't know whom Romesh knew'.)

Another reason for leaving the issue open is that the as yet unstudied grammar of antecedents[17] will probably tighten to some degree the constraints which the sentence grammar

[17]At the PhD defence (between my September 1979 first draft and February 1980 final draft), committee member Ray Dougherty asked me why a thesis on questions offers no comments the question-answer relation. The mainstream accounts of questions then available in formal semantics took it that a question is a partitioned set of possible answers; I did not see a way to unify that enterprise with a theory that takes the wh connection between relatives and questions seriously; so I steered clear of that insufficiently fecund line of inquiry.

imposes on well-formedness of relative-antecedent linkages, even without labelling particular rules as obligatory. Perhaps the most immediately obvious facts about antecedents which I should mention, without going into detail, are that *Emon* 'of this sort', the antecedent marker which cropped up in section 8.5 (example (54)), is typically obliged to serve as a binder wherever it can do so, and that Distal Demonstrative forms are usually barred from functioning as antecedents to relative phrases, in contrast to Sequent forms which are antecedents par excellence and Proximal Demonstrative forms (like *Emon*) which also do occasional duty in this role.

When launching the Grammatical Interpretation Hypothesis I almost promised that the analysis to be proposed under the aegis of GIH would be able to account for the availability of equational as well as unique-existential readings for sentences containing relative clauses; recall the beginning of section 8.8. Now that I have constructed a GIH semantics for relative clauses, one may well ask whether I have kept my word. It is difficult to answer, given the absence of rules for mapping from my variable-binding level into full-fledged logical representations (if such mapping is what is desirable). But the route from bound variables to a Mostowski-Tennant equationally quantified translation does seem pretty straightforward to me. The unique-existential quantification is harder to reach from here, unless one relies partly on grammatical cues (like number and 'definiteness' connotations of a NUM′ structure which follows rather than precedes its N) and partly on pragmatic factors. Perhaps a good part of the solution to this problem will come out of study of the 'definiteness' aspect of the relative clause construction itself; see chapter 9 for a programme for such research.

Another question which I have left unanswered is whether the results just obtained carry over smoothly to Right Relatives, although all my examples have been of Left Relatives. The answer, paradoxically, is that (70) works so smoothly for all sorts of Relatives that it overgenerates in the case of Right Relatives, and thus fails to predict that Right Relative Clauses must have only one relative phrase per clause, unlike Left Relatives. One response to this problem is to perform a tactical retreat, to revise (70) to the more complex (and therefore less satisfactory) rule (71), and to suppose that a similar rule must be used for Right Relatives but that that rule will only work when the Relative Clause has only one relative phrase in it.

(71) Relative Binding (final version)

S.D.: an S′ whose COMP is –INT and with which a construal $((y_1, y_2, \ldots, y_n), x)$ is associated, embedded in a larger S′

S.C: make it so that each variable is bound by at least one antecedent to the right of the embedded S′ and any single antecedent in the matrix S′ directly binds at most one variable

I will provisionally take the stand stated at (71), for the purposes of this dissertation. Someone working on Right Relatives may be more interested in exploring another option which suggests itself. Instead of placing the burden of accounting for the one-relative-phrase-only property of right relatives on the binding rule, one might argue that, just as Left Relatives correspond to

---

Instead, I hoped that the relative-antecedent bond could be explored more rigorously. As far as I know, not much of significance has been done on these fronts since 1979, not even in semantics.

Complement Clauses (which contain the real Complementizer *je*), so also Right Relatives correspond to the Pseudo-Coordinate (or Pseudo-Complement) Clauses with the Conjunction *je* which are discussed in chapter 9. The latter correspondence may be made to follow from some construal-like rule which links the unique relative phrase in a Right Relative Clause to a necessarily empty Conjunction node marked +OPEN –INT. Such a course would have the merit of explaining at least the obligatory extraposition (or what appears as such) of Right Relative Clauses; the explanation, on those assumptions, would be that, in S CONJ S, the CONJ S sequence occurs after the first S, in an 'extraposed' position. Besides, if the appearance of relative preposing in Right Relative Clauses is at all well-founded, that too may have to do with the CONJ node; perhaps one has a choice between just construing the relative phrase with an empty CONJ node and actually moving the phrase into the CONJ position.

If one is able to elaborate and defend such an account, one will not have to complicate (70) to (71), the difference between left and right relatives now being attributed to their different structures.

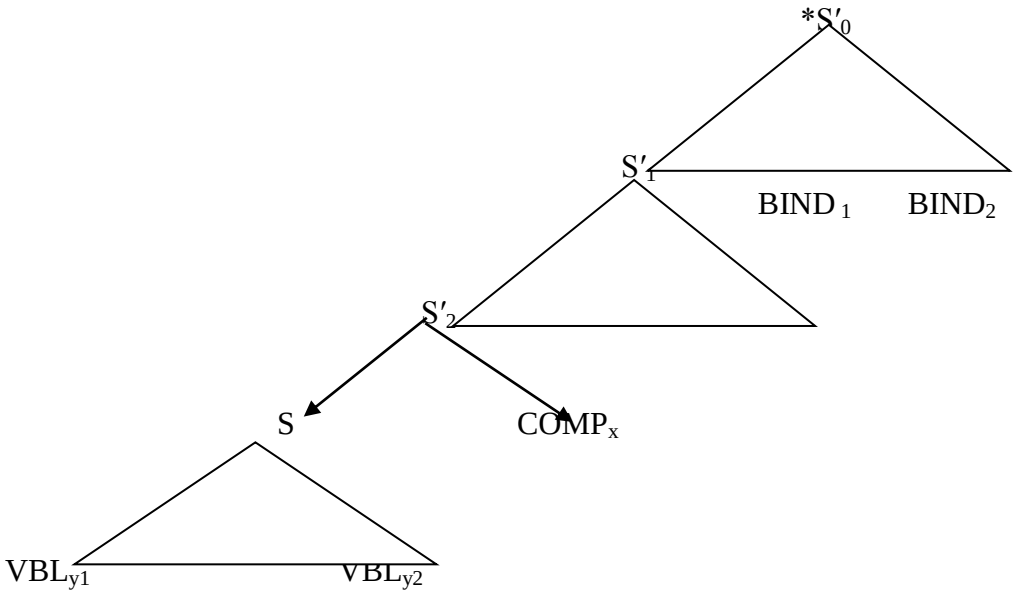
Another option open to a serious student of Right Relatives is to attribute the difference to the different positions of the two sorts of relative clauses. Mark Baltin has pointed out to me that his (1978) schema for construing extraposed modifiers with their heads leads to the expectation that Right Relatives will have fewer capabilities than Left Relatives. As mentioned in section 8.6, Baltin's work in its present form does not directly apply, since he only studies extraposed cases, while the facts of Hindi show that even non-extraposed Right Relatives (in a language which is otherwise like Bangla) are incapable of containing more than one relative phrase. But it may prove possible to extend Baltin's result to all right modifiers. In that case one would, as in the possible CONJ-focused alternative account mentioned above, leave (70) as is and allow universal principles to predict that only Right Relatives will impose a ceiling of one on the number of relative phrases per clause.

8.12 Simplexity: A Statement

One entailment of my theory of relative structures – an entailment that may be viewed as the main contribution made by the present dissertation – deserves separate study, partly because of its intrinsic interest, and partly because it provides a clear-cut means of comparing (for adequacy) the present proposal with natural alternatives. I call this entailment Simplexity. The following exposition builds up towards a statement of Simplexity. The statement comes at the end of this short section (and is best read with the Appendix to the chapter, added in 2020). Section 8.13 distinguishes real from apparent counterexamples to Simplexity.

Since rule (71) operates on a construed S′ embedded in a matrix S′, this rule always takes an operand where no S′ node simultaneously is dominated by the matrix and dominates the embedded S′. To see this, consider a tree of the 'wrong' form, with S'_1 illegitimately simultaneously being dominated by the matrix S′ and dominating the embedded S′.

(72) The Pre-Anti-Simplex Tree

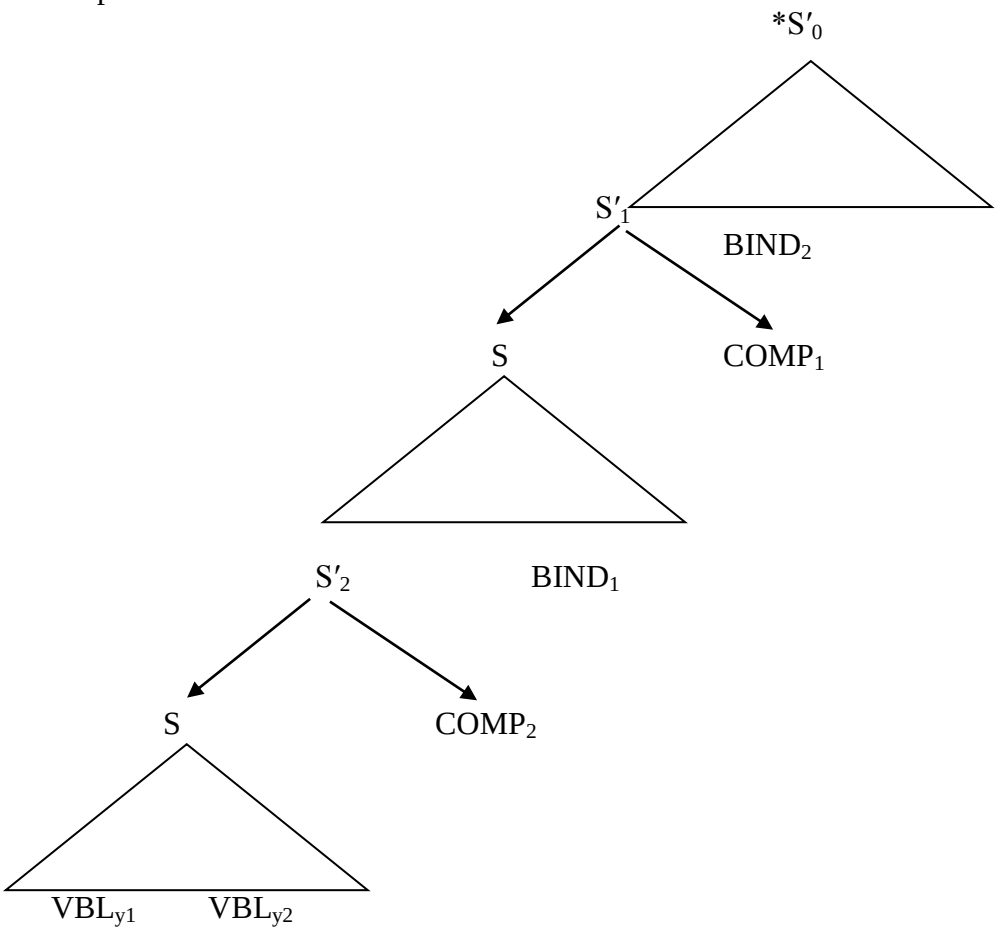


The only way in which structure (72) can come out well-formed is if S'_1 happens to contain phrases eligible to bind into S'_2; if it does, those phrases can take care of the needs of the variables in S'_2, and therefore $BIND_1$ and $BIND_2$, not being needed for S'_2, can either fend for themselves as free anaphors or be used to bind relatives (if any) in S'_1. However, if S'_1 has nothing to offer to bind the variables in S'_2, it would be illegal to attempt to relate $BIND_1$ and $BIND_2$ to y_1 and y_2. Such an attempt would violate the domain-under-domain principle which is built into the derivational cycles (the transformational cycle, the phonological cycle, etc.) of explicit generative grammars and was a proviso to the original statement of the A-over-A condition in *Language and Mind*. In the absence of any antecedent-relative relation tying something in S'_1 (but outside S'_2) to parts of S'_2, nothing else will semantically relate S'_2 to the rest of S'_1 – a state of affairs which, for reasons that seem self-evident and are therefore rather hard to put into words, is intolerable.

The impossibility of variable bindings of the sort shown in (72) is a consequence of general principles acting in concert. This prohibition rather resembles the conception that Bresnan at one time had of Subjacency as a principle applicable to deletion as well as movement. The difference is that, where Subjacency says that positions A and B may be related both if they are clause-mates and if one bounding node boundary separates them, the prohibition we have just seen says that positions A and B may be related only if an S′ boundary (and no more than one S′ boundary) separates them. It would be rather surprising if the rule relating relative phrases to their antecedents were to observe Subjacency. It does not observe the very widely applicable Left Branch Constraint, as one sees when one looks at the relation between *the man* and *whose* in *the man whose sister Bill married*, or, in Bangla, *bil je lokTar bonke biye koreche Se* "Bill which man-item-GEN sister-OBJ marriage has-done he". Recall that one of the motivations for Chomsky's postulation of subjacency was that the notion of subjacency would parsimoniously explain facts formerly seen in terms of more complicated constraints like the Left Branch Constraint.

Consider now the following slightly different tree.

(73) The Anti-Simplex Tree



One sentence which exemplifies the schema of the Anti-Simplex Tree is the ungrammatical sentence **je chatrora je hOsTelTate thake SeTate Dak bibhager kormira jaY na tara bideSi* "which students which dormitory-item-LOC live that-item-LOC postal department-GEN workers go not they foreign" 'Students such that the dormitory where they live does not receive postal deliveries are foreign'[18] (to reproduce the ill-formedness in one's gloss, one might try the following: 'Students who_i the dormitory in $which_j$ t_i live t_j gets no postal deliveries are foreign'). In general, manifestations of (73) are ill-formed. Let us consider how our rules would describe the ungrammaticality of (73), in order to understand the system better.

If $COMP_2$ is associated with x, then S'_2 is the relative clause subject to Relative Binding, and, as shown above, S'_1 must then be the matrix, in which case $BIND_1$ has to bind x in order to ensure that no variable is left unbound. $BIND_1$ thus subbinds y_1 and y_2. In our example, this step would yield the nonsensical result that *tar* 'its' subbinds *je chatrora* "which students" and *je hOsTelTa* "which dormitory".

If $COMP_1$ were to be associated with x – just for the sake of argument, although the

[18]On being told the facts, with a gloss of this sort, Steve Cushing (p.c. in 1978) exclaimed: 'You guys talk Montague Grammar!'

construal rule which builds the set of variables wouldn't allow this in the first place – then S'_1 would become the relative clause subject to relative binding, and S'_0 would be the matrix, in which case $BIND_2$ would have to bind x in order to ensure that no variable is left unbound. $BIND_2$ would thus subbind y_1 and y_2. In our example this step would yield the nonsensical result that *tara* "they" subbinds *je chatrora* "which students" and *je hOsTelTa* "which dormitory".

Thus, there is no way for the present system of rules to generate (73) with the pairings shown. I will call this entailment of the system **Simplexity**; the trees which this entailment outlaws will be called Anti-Simplex strings, or Simplexity violations.

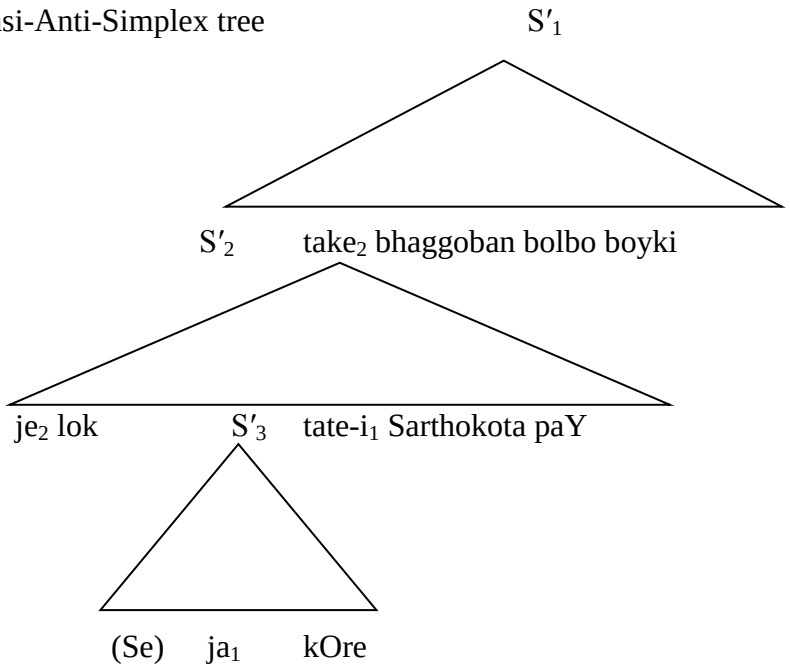
8.13 Well-formed quasi-anti-simplex Strings

There is an immediate difficulty. Any native speaker of Bangla reading this study is liable to come up with well-formed sentences like *je lok ja kOre tate-i Sarthokota paY take bhaggoban bolbo boyki* "which person what does that-LOC-EMP fulfillment gets him/her fortunate will-call of-course" 'I/we will, of course, call the person fortunate who feels fulfilled in whatever (s)he does' and point them out as counterexamples to the Simplexity entailment and therefore to the theory which entails it. Such an objector may even flag the fact[19] that these counterexamples have a distinctive intonation contour (in the sentence just mentioned, /ja/ carries focal pitch) – which, unfortunately, does not help at this stage of our ignorance, since we are very far from having a characterization of the Bangla intonation system. I am forced to focus on the bald syntax of these sentences.

As far as I have been able to ascertain, all such apparent counterexamples, on careful case-by-case study, turn out to call for a syntactic structure which does not violate Simplexity. The example just given, for example, seems to be derived by deletion of *Se* from (or to be related — by allomorphy of *Ø* and *Se* or by interpretation of *Ø* as in effect *Se* — to) the synonymous (and, for what it is worth, similarly intoned) sentence *je lok Se ja kOre tate-i Sarthokota paY take bhaggoban bolbo boyki* "which person (s)he what does that-LOC-EMP fulfillment gets him/her fortunate will-call of-course". And, like this paraphrase, the example calls for assignment of the following structure.

[19] Full disclosure: this is a 2020 interpolation. At this stage of the game, this remark may even help someone seeking to revisit the issues, since our phonologists do understand Bangla intonation a lot better today than they did in 1979.

(74) A typical Quasi-Anti-Simplex tree



This structure, in which *Se* is freely anaphoric to *je lok*, clearly does not violate Simplexity. Each relative phrase is within 'reach' of its antecedent. See the Appendix for more.

8.14 Simplexity and Constituent Questions

I could follow up the above exposition of Simplexity by showing, in detail, that J. Chattopadhyay's analysis of relative clauses as it stands does not entail Simplexity as a generalization. However, I will not try to show this, firstly because it is already clear that that analysis is inadequate as it stands, and secondly because surely Chattopadhyay would, when faced with new data, modify her theory to accommodate them, and, not knowing what modifications she would resort to, I cannot argue against her position in a vacuum. I will therefore explore instead the possibility of extending the formulation of Simplexity so that it covers constituent questions as well as relative clauses.

Simplexity pertains to binding. In order to extend simplexity to questions we need to redescribe their semantics in terms of binding.

I will provide two such redescriptions. The first one, although less good as an approximation, simplifies the exposition about Simplexity. The second one is the closest approximation to the truth in this area that I have been able to formulate so far. It is surely not the last word.

My original description of the semantics of constituent questions was in terms of 'no binding', in terms of leaving variables 'unbound'. I now offer a redescription in terms of 'binding by nothing' that renders the treatment of questions analogous to the operation of the Relative Binding rule. Just as the relative phrases in the clause *je chele je puroSkar caY* "which boy which prize wants" are associated with variables y_1 and y_2 which get bound by the antecedents *Sey chele* "that boy" and *Sey puroSkar* "that prize", we shall imagine that the

variables y_1 and y_2 associated with the interrogative phrases *kon chele* "which? boy" and *kon puroSkar* "which? prize" in the constituent question *kon chele kon puroSkar caY?* "which? boy which? prize wants" get bound by antecedents, except that these antecedents are 'non-existent' in the sense of non-presence in the tree – they are concatenation elements, marked e_1 and e_2 respectively, located as specific absences in the phrase-marker.

It takes no great effort to make such elements available to one's formal imagination. Recall the discussion of adjunction and classical pruning in chapter 4, where it was mentioned that one of the standard assumptions of generative grammar is that, for any node A, $[_A A\ e] = A$; this is also true for a sequence of e's. Thus, the standard assumptions would in any case underwrite the statement that the S′ *kon chele kon puroSkar caY* can be assigned the phrase-marker $[_{S'}[_{S'}$ kon chele kon puroSkar caY] e e], even if we had no use for the availability of this alternative form of the phrase-marker. It so happens that my redescription of questions, (75)[20], does have some use for it.

(75) Interrogative Binding (first approximation)
S.D.: an S′ with which a construal $((y_1, y_2, \ldots, y_n), x)$ is associated, whose COMP is +INT, and which is in the environment $[_{S'}$ ___$e_1, \ldots, e_m]$
S.C.: make it so that each variable is bound by at least one e_i and any single e_i directly binds at most one variable

Many readers will feel that this rule does not materially alter my analysis of constituent questions as containing unbound variables, that binding by nothing is a notational variant of no binding. Not being sufficiently well versed in the philosophy of nothing (nihilology?) to answer this objection frontally, I merely point out that the redescription offered makes it possible to imagine the following phrase-marker, and that the sentence so structured happens to be ill-formed, as my theory predicts; and I point out that, without (75), there is no obvious way to explain the ungrammaticality of sentence (76):

(76) *$[_{S'0}\ [_{S'1}\ [_{S'2}\ [_{S'3}\ [_{S3}$ kon$_{y1}$ chele kon$_{y2}$ puroSkar caY] COMP$_x$] e_1] ram jane] e_2]
"which? boy which? prize wants Ram knows"
intended reading: *'Which boy does Ram know wants which prize?'

Given (75), as well as the convention of Overbinding and the Simplexity entailment of my analysis, I can of course carry out a straightforward extension of my theory to cover the ill-formedness of (76); for (76) manifests the forbidden Anti-Simplex tree schema. The possibility of thus explaining the ill-formedness of (76) provides an argument supporting rule (75) and indirectly supporting other elements of my theory which enable this binding rule to have the effect it has.

I must admit, though, that rule (75) overgenerates: it allows for possibilities that are never utilized. For example, it predicts that – parallel to the ambiguity of *ram jake ja Sikhiyechilo Ekhon Se ta bhulte caY*, which may mean 'the person Ram taught it to now wants

[20] A 2020 note: I may need to stress the fact that this dissertation was written without access to the Pisa lectures and therefore 'before' the theta-criterion. The wording of (75) stemmed from logical considerations, not from a response to the REST tradition, which I had distanced myself from.

to forget what Ram taught him/her', or 'Ram now wants to forget about having taught what he did to the person he taught it to', or 'person z now wants to forget about Ram having taught what he did to the person he taught it to' – the sentence *ram kake ki Sikhiyechilo?* "Ram whom? what? taught" 'What did Ram teach to whom?' should be ambiguous between two readings: one in which two nothings bind the variables associated with *kake* "whom?" and *ki* "what?" respectively, and another reading that has a single nothing binding the variable associated with the COMP of the S′ and thereby subbinding y_1 and y_2. (There is no third reading, since there is no 'second nothing' that might sit around and choose whether to be freely anaphoric to *ram* or to be freely anaphoric to someone else in the discourse or in the world. Nothings, dependent as they are on the semantic rule, do not have the privilege of free anaphora.) However, no such ambiguity is available, unless my semantic judgment is too blunt to detect it. As far as I am concerned, the two readings prediction is false, and the theory that makes it must be revised from (75) to (77).

(77) Interrogative Binding (second approximation)
S.D.: an S′ whose COMP is +INT and with which a construal $((y_1, y_2, \ldots, y_n), x)$ is associated
S.C.: enlarge the construal to include an e binding x

Philosophers will notice that (77), unlike (75), no longer offends the principle of identity of indiscernibles, which is occasionally called Leibniz's Law.

(77) empirically differs from (75) in two respects. The first difference, immediately motivated by the counterexample cited, is that I now permit only Split control for the interrogative case (as opposed to the relative case) and thus no longer make the incorrect prediction that every multiple question should be multiply ambiguous. Since this move has done away with the need to use the potential presence of several nothings (rather than just one nothing), I felt it would be advisable to shift from a syntactic basis for the binding element to a semantic basis. Hence – and this is the second difference – the empty binder in the new rule, unlike the nothings of (75) which were made available by the standard assumptions of generative grammar, is explicitly put into the semantic structure by the rule of binding. The empirical effect of this is that we are no longer making the claim (in case it was a 'claim' with descriptive consequences) that S′ plus one or more nothings is a syntactic node!

If we adopt (77), the new analogue to (76) is an Anti-Simplex structure which doesn't have all the extra clause nodes but does have two nothings associated with the construals of the two clauses, illegitimately, since (77) will only target the clause S'_3. Thus, we still more or less retain the extension of Simplexity to cover questions as well, although the isomorphism with the relative case is formally less clear than it would have been under the analysis based on (75). Some readers, perhaps, will report that they perceive an ambiguity (or even several ambiguities) in *kon chele kon puroSkar caY?*; for their 'dialect' we can retain (75), with its clearer similarity to (70).

Although it is obviously possible to conflate Relative Binding and Interrogative Binding into a single rule or rule schema, I leave the question of the form of such conflation open. And I will not even raise the question of formal unification of one or both of these rules with rules which handle the semantics of the clause types of chapter 9.

We are ready now to attempt a description of Complement Clauses and Yes-No Questions, and, on the basis of this description, to formulate the principal thesis of this

investigation with some exactness.

Appendix (added in 2020): Apparent counterexamples to Simplexity: further remarks

Readers today will find it hard to extricate the 1980 formulation of Simplexity from the inappropriately configured verbiage surrounding it. It is best to offer a fresh formulation:

(73S) **The Simplexity Condition**

> Consider a sentence containing a relative clause RC featuring two or more relative phrases. In such a sentence, all the sequent phrases coreferential to these relative phrases occur in one and only one sequent clause SC. A pairing of relatives with sequents that exhibits this property is called a simplex pairing. **All correlative structures are simplex.**

My thesis mentioned all too briefly the crucial fact that, if one wishes to state the simplexity condition more precisely, one must deal with apparent counterexamples like (74), which I termed 'quasi-anti-simplex sentences'. It is possible that workers today will wish to explore the range of such apparent counterexamples more rigorously. The purpose of the following preliminary remarks is to encourage such inquiry; the immediate goal is to delineate the range accessible to structure (74); more is said, below, on goals that lie beyond that point.

(78) is parallel to (74), except that (78) harbours *two* relatives, not just one, in S'_3.

(78) [[je SaSok [(Se) *jOkhoni ja* caY] tOkhoni tay peYe jaY]take bOle ekocchOtro SOmraT]
"which ruler (s/he) *when-Emph what* wants then-Emph that get Aux him/her one-calls omnipotent emperor"
'A ruler who gets whatever they want whenever they want it is an omnipotent emperor'

In (79), we ring the changes by fronting an object rather than a subject; in (80), we front an adjunct; in (79) and (80), the reason we do not show an optional sequent phrase parallel to the /Se/ of (74) and (78) is that such a sequent phrase cannot occur – the construction has properties suggesting what later work would call parasitic gaps, a topic orthogonal to our concerns here. These sentences, like (74), are built around generic phrases:

(79) [[je porikkha [je-i deY] Se-i paS kOre] SeTa adow porikkha hObe ki kore?]
"which exam who-Emph writes s/he-Emph pass Aux that-item at-all exam is how Prt"
'If whoever writes this exam passes it, how is it an exam at all?'

(80) [[je ghOre [jara-i rat kaTaY] tara-i mara jaY] take golmele to bolbe-i loke]
"which room-Loc who-Pl-Emph night spends they-Emph die Aux it-Obj suspicious Prt will-call-Emph people"
'If whoever spends the night in a room dies, people will naturally view that room with suspicion'

Definite descriptions, however, are also possible – the generics vs definite reference contrast is yet another red herring:

(81) [[je chatrora [gronthagare (tara) jeSOb boy khuMjechilo] Segulo paY ni] tara jEno amar SOngge dEkha kOre]
"which students library-Loc (they) which-Pl book sought those found not they Prt me with appointment make"
'The students who didn't find the library books they had looked for should see me'

(82) [[je khodder [(Se) je dokane cirokal boy kene] Sekhane-o agun dam dekhche] Se to boy kena bOndho korbe-i]
"which customer (s/he) which shop-Loc always book buys there-Emph fire price finds s/he Prt book buying stop will-Emph"
'A customer who finds prices skyrocketing even at the shop where s/he always buys books will naturally stop buying books'

How, then, do we narrow down what it is that prevents the quasi-anti-simplex structure from mimicking the truly anti-simplex tree (73), whose prototypical example given earlier is repeated here, for expository convenience, as (83)?

(83) **je chatrora je hOsTelTate thake SeTate Dak bibhager kormira jaY na tara bideSi*
"which students which dormitory-item-LOC live that-item-LOC postal department-GEN workers go not they foreign"
'Students such that the dormitory where they live does not receive postal deliveries are foreign'

If we try to place the words of (83) in the quasi-anti-simplex tree, we fail, for reasons that become apparent when one looks closely at the phrase structure of a case in point, (84):

(84) *[$_{S'1}$ [$_{S'2}$ *je chatrora* [$_{S'3}$ *(tara) je hOsTelTate thake*] *SeTate Dak bibhager kormira jaY na*] *tara bideSi*]
"[[which students [(they) which dormitory-item-LOC live] that-item-LOC postal department-GEN workers go not] they foreign]"
'Students such that the dormitory where they live does not receive postal deliveries are foreign'

The syntactic factor that stars (84) is, obviously, the non-uniqueness of candidates for the subject position of the verb /jaY na/ 'don't go' in sentence S'_2. There is no way for both /je chatrora/ "which students" and /Dak bibhager kormira/ "postal department-GEN workers" to occupy the subject slot of that verb; Bangla has no multiple subject construction.

Further questions arise when one moves out of the argument structure of the clause and starts exploring anti-simplex and quasi-anti-simplex examples keeping in view additional issues, involving what since the eighties has been called A-bar positions. I am not pushing this appendix to that point; the exercise would take us too far afield.

The late Alice Davison made a potentially significant point in the eighties. She noted that well-formed multiple relatives like (36) [repeated below as (85)], which fully conform to the expectations embodied in the simplexity condition, have the consistent property that every sentence of this type is synonymous to some 'donkey sentence' (i.e. some sentence that exhibits

the format of *Every man who owns a donkey beats it*) – in this case, (86) is the 'donkey sentence' synonymous to (85):

(85) je chele je puroSkar caY Sey chele Sey puroSkar pabe
"which boy which prize wants that boy that prize will-get"
'The boy who wants it will get the prize he wants'

(86) je chele EkTa puroSkar caY Sey chele Sey puroSkar pabe
"which boy a prize wants that boy that prize will-get"
'The boy who wants a prize will get it'

For the full significance of Davison's observation to become clear, syntactic and semantic study will have to focus on at least the following phenomena.

Fact One: Donkey sentences do not observe simplexity. (87), a correlative structure that violates simplexity and is ill-formed, and (88), the same structure with a donkey sentence replacing the multiple relative clause and coming out well-formed, constitute a minimal pair:

(87) *je jOnojati je gachke pobitro bole meneche Se jati rukhe daMRabe jodi kew Sey gach keTe tar kaTh bikri kOre
"which tribe which tree-OBJ sacred as has-recognized that tribe fiercely resist-FUT if Anybody that tree having-cut its wood sell does"

(88) je jOnojati EkTa gachke pobitro bole meneche Se jati rukhe daMRabe jodi kew Sey gach keTe tar kaTh bikri kOre
"which tribe one-CLA tree-OBJ sacred as has-recognized that tribe fiercely resist-FUT if anybody that tree having-cut its wood sell does"
'A tribe that has recognized a particular tree as sacred will put up fierce resistance if anybody cuts that tree and sells its wood'

Fact Two: If one uses an Emphasizer /i/ to relieve a relative phrase from its normal duties and turn it into a quantified phrase, one replicates the behaviour of a donkey sentence, as in (89):

(89) ekoda jeSOb jOnojati jOto jaYgaY jOto gachke-i pobitro bole mene thakuk na kEno, aj SeSOb jatir bohu bOngSodhOr kono apotti korbe na jodi bEbSayira SeSOb gach keTe tar kaTh bikri kOre
"ages-ago which-all tribe how-many place-LOC how-many tree-OBJ-EMP sacred as recognize-and AUX Prt Prt, today that tribe-GEN many descendant any objection will-do Neg if tradesmen that-all tree cut-and its wood sell do"
'Whichever tribes may have recognized however many trees as sacred in however many places, today many of their descendants will not object if tradesmen cut those trees and sell the wood'

The Emphasizer /i/ in (89) certainly turns /jOto gachke-i/ "how-many tree-OBJ-EMP" 'however many trees' into a quantified expression; whether the effect also extends to /jOto jaYgaY/ remains to be determined, and described carefully.

In connection with the task of exploring further the implications of Davison's

observation, I would also like to point to the pertinence of the mechanism of 'absorption' that was introduced in the early eighties to handle the fact that multiple interrogatives get their complementizer to handle all the constituent interrogation threads in one clausal cluster, and equivalent mechanisms in later work.

Suppose we were to extend absorption to cover multiple relatives as well. Once we make that move, firstly, the scope of simplexity indicated above (the fact that it is confined to true relative clauses in which no emphasizer converts a relative phrase into a quantified expression) follows as a theorem, and secondly, the extended absorption account becomes a notational variant of simplexity. Such a variant, to be sure, would prioritize the grammar of questions over the grammar of relatives. A theory with this property would still be fine if it proved able to deal with multiple relative structures. The record shows that workers who have followed that path have not yet come up with a successful account.

The main stumbling block, one must realize, is that we are still waiting for a comprehensive account of 'free relatives' in English-type languages. Othello says: "Who steals my purse steals trash" (Shakespeare's *Othello*, Act III, Scene 3, line 162, please don't force me to add it to the bibliography). This is no longer a well-formed sentence in today's English, whereas "What is good enough for you is good enough for me" is fine. The relevant distinction between *who* and *what* has never been made clear in any current theory of syntax. More importantly, we do not have an account from which it follows that English does not permit multiple free relatives. If English had in fact allowed "Who wants what will get it", this sentence would have meant 'For every x and y such that x wants y, x will get y'.

Whether a successful account will start with interrogatives and 'absorb' relatives, or the other way round, is not the issue. The point is to come up with **some** account that makes sense of the fact that Bangla-type languages harbour multiple relatives and that English-type languages do not. My 1980 project was to come up with a descriptively adequate analysis of Bangla. In the decades that have followed, scholars who claim to have elaborated 'universal grammar' in ways that can cover all known significant syntactic facts have yet to take the properties of Bangla and of English in this domain and place them on a coherent map.

This is why further digging is required (and while digging, it's important to be thorough, e.g. check with many speakers to see if replacing /EkTa/ 'a' with /kono EkTa/ 'some' in (86) makes any difference) in order to understand better what makes the apparent violations of simplexity tick, why there are no actual violations, and, beyond this point, what the comparative grammar of these constructions looks like. Before you universalize, you need to compare, rigorously.

Chapter 9

COMPLEMENT CLAUSES AND YES-NO QUESTIONS; THE THESIS

9.0 Strategy and tactics

Roger Garaudy says somewhere that each step of intellectual activity is both absolute and relative: absolute with respect to the foregoing work it sums up and organizes, relative with respect to those things which it leaves visibly undone and thus opens up for further inquiry. Whatever general value Garaudy's remark may or may not have, it is certainly pertinent to the work in this chapter. Chapter 9 brings the whole investigation to a head by stating the principal claims of the dissertation and discussing the scope and limits of these claims. This chapter also formulates new questions which the research reported here makes it possible and desirable to pursue. As a prelude to such pursuit, chapter 10 shall relate this dissertation to the Chomsky-Bresnan debate.

In chapter 8 I have discussed DET-Open Clauses, i.e. clauses containing one or more phrases with an Open Word in DET position. As noted, for well-formedness a DET-Open Clause must have an Open COMP. This chapter begins by studying COMP-Open Clauses, clauses with an Open Word in COMP position, which, for reasons that will become clear, therefore are not allowed to contain phrases with Open Determiners. The two major classes of COMP-Open Clauses – Complement Clauses and Yes-No Questions – are dealt with in that order.

Section 9.6 encapsulates the findings of the foregoing investigation into a principal thesis and elaborates this thesis on the planes of general and particular linguistics. A brief discussion of further research made possible by the work done here follows.

9.1 (Declarative) Complement Clauses (with /je/)

The emphasized portion in each of the following sentences might be called a 'declarative Complement Clause with *je*' or, to use Jayanti Chattopadhyay's standard and convenient abbreviation, a 'Complement Clause'.

(1-S) *ram je istOpha diyeche* ami Se kOtha jani
"Ram that resignation has-given I that matter know"
'I know (the fact) that Ram has resigned'

(1-C) ami jani *je ram istOpha diyeche*
"I know that Ram resignation has-given"
'I know that Ram has resigned'

However, I find such terminological conflation misleading. I am using C and S to flag my claim that (1-C) is a coordinate structure in which *je* serves as a Conjunction, whereas (1-S) is a subordinating structure in which *je* functions as a Complementizer to the S *ram istOpha diyeche*. Corresponding to this structural difference between the C and S constructions, we find

several behavioural differences.

First, the C but not the S structure permits the clause with *je* to express a result or effect.

(2-C) Emon EkTa kOtha uTheche je dOler SOkoleri kOnTho SOrOb
"such a matter has-arisen that group-GEN everyone-GEN voice animated"
'Such a topic has been raised that everybody in the group is speaking animatedly'

(2-S) *dOler SOkoleri kOnTho je SOrOb Emon EkTa kOtha uTheche
"group-GEN everyone-GEN voice that animated such a matter has-arisen"

Sentence (2-C) is a quote from Bimal Kar's novel *Songgini* (Kar 1969:11).

Second, in the C but not the S structure the clause preceding *je* may deploy such predicates as *dharona* or *biSSaS*.

(3-C) amar biSSaS/ dharona je Somnath aj mOhoRa debe
"(it is) my belief/ understanding that Somnath today rehearsal will-give"
'I believe/understand that Somnath will rehearse today'

(3-S) *Somnath je aj mOhoRa debe amar ta/ Sey biSSaS/dharona
"Somnath that today rehearsal will-give my that belief/ understanding"

Third, the C but not the S form lets the clause following *je* serve as a contextual restrictor to a question.

(4-C) gandhi ke je amra taMr kOtha Sunbo?
"Gandhi who? that we his word should-listen"
'Who is Gandhi that we should listen to him?'

(4-S) *amra je taMr kOtha Sunbo gandhi Sey/Emon ke?
"we that his word should-listen Gandhi that/such who?"

Fourth, the C but not the S structure allows the clause following *je* to be built around an imperative V.

(5-C) ram caY je SEm kaj koruk
"Ram wants that Shyam work do"
'Ram wants Shyam to work'

(5-S) *SEm je kaj koruk ram ta caY
"Shyam that work do Ram that wants"

These and other behavioural differences between the C and the S structure rule out any straightforward transformational derivation of C from S (as I had expected to postulate when I set out to do the research) or of S from C (as Chattopadhyay (1976a, b) postulates). It is unreasonable to try to build into the statement of a syntactic rule such things as sensitivity to the mood of the matrix verb, sensitivity to locutions of belief, and other sensitivities

corresponding to each of the intricate differences between the two structures. We must base-generate the two constructions separately and interrelate them by means of semantic rules.

It would have been reasonable even if we postulated a transformation, and it is necessary since we are not postulating one, to regard the C and S structures as distinct and therefore to give them different names. We cannot afford to call them both 'Complement Clause' structures. The choice is surely obvious. Surely it is *ram je istOpha diyeche* in (1-S), with the true Complementizer *je*, which must be called a 'Complement Clause'. Let us call *je ram istOpha diyeche* in (1-C) a '(declarative) Pseudo-Complement Clause'. I will leave Pseudo-Complement Clauses for future investigators to tackle. In this volume I will concentrate on true Complement Clauses – those in which the COMP, being a true COMP, can undergo COMP Preposing and other COMP-targeting processes. Although *ram je istOpha diyeche* is a relatively short clause, even this clause can be preposed:

(6) ram istOpha diyeche je (deep structure)

(7) ram istOpha je diyeche

In contrast, the Pseudo-Complement Clause, of course, not having a true COMP (see the discussion of Conjunctionization in chapter 6), does not allow preposing; conjunctions do not prepose. No permutation of (1-C) obtained by moving *je* is well-formed.

An alternative designation for pseudo-complement clauses – 'pseudo-coordinate clauses' – might in fact make more sense. There are two pseudo things about them – two things they do that 'genuine' coordinate clauses, those which display a full range of coordinate clause properties, don't do. For one thing, they semantically function as predication relata, as if they were subordinate clauses. For another, their intonation is very different from that of typical coordinate clauses – the clause break comes after rather than before the conjunction and does not show standard terminal contour the way bona fide coordinate clauses do. Yet, there seems to be no way around regarding these clauses as coordinate clauses, since their introducer *je* fails to co-occur with a bona fide conjunction: *ami jani je S ebong je S* "I know that S and that S" (right way to say it: *ami jani je S ebong S* "I know that S and S"). Future investigators will no doubt settle the terminological issue, in the context of findings that we are in no position to anticipate. We shall revisit the terminology question in section 9.5, in connection with interrogative pseudo-coordinate clauses.

9.2 Complement clauses and left relative clauses

In my analysis of all relatives, and in particular of left relatives (which are under specific scrutiny in this dissertation), I have assumed that the +OPEN –INT COMP node of relative clauses is empty and that a semantic rule construes this empty COMP with the relative phrase in the clause.

We are now looking at complement clauses, where the COMP node is not empty but dominates the word *je*, which is itself (see chapter 7) a Relative Word. So, we now claim that in a complement clause, as in a left relative clause, the COMP node is +OPEN and –INT, but that the COMP position is, moreover, filled. This claim has consequences for (A) construal and (B) binding.

(A) The consequence for construal is that complement clauses differ from relative clauses in the following respect. In a clause where the COMP position is filled by a Relative Word, COMP is not available for construal with Relative Determiners. Therefore, a complement clause, whose COMP position is filled by a Relative Word, depends for its well-formedness and interpretability on not containing any Relative Phrases (phrases with a relative determiner in them). A clause whose COMP is +OPEN –INT must either have relative phrases and an unfilled COMP (and thus be a relative clause) or have a filled COMP and no relative phrases (and thus be a complement clause). In the former case, there will be construal of the relative phrases with the COMP. In the latter case, no construal will occur. Strictly speaking, the expressions 'filled COMP' and 'unfilled COMP' three sentences ago should have read 'COMP filled/unfilled *by a Relative Word*'. Thus in some dialects, relative clauses may contain the overt Complementizer *naki* or the overt Complementizer *kina*, which are not Relative Words and which seem not to make such a semantic contribution as would hamper the process that construes relative phrases in a clause whose COMP node is –INT. In my speech, the COMP node in a relative clause is always literally empty. But, in order to assume no greater difference between my dialect and the other one than is clearly motivated by the data, I will assume that the INT feature is directly associated with COMP (rather than have COMP dominating an +OPEN daughter which takes responsibility for the INT feature) in order to accommodate the fact that *naki* and *kina*, which are not Open Words, do not obstruct construal of relative phrases with the COMP in the other dialect; there is no reason why my dialect should not also have COMP sometimes directly bearing the ±INT feature. The fact that I was only **tacitly** supposing all along that the empty COMP to be construed and/or bound would always be +OPEN now pays off; I can now simply insist that my rules be taken literally, that I never said that the COMP in a construal had to be marked +OPEN.

(B) The consequence for binding is that complement clauses resemble relative clauses in the following respect. Some binder outside the clause will directly bind the filled COMP of a complement clause (the process of binding will stop there, since there are no construed relative phrases to subbind) exactly as some binder outside the clause can directly bind the unfilled COMP of a relative clause (in which case the binder also must subbind any relative phrase(s) construed therewith which are not otherwise bound).

These consequences entail the prediction that, say, (8) and (9) should be well-formed and (10) ill-formed – a prediction that is borne out by the facts, as we see.

(8) [[ram SEmer SOngge [$_{COMP_i}$ je] kOtha bolchilo] jodu Se$_i$ kOtha Suneche]
"Ram Shyam-GEN with that word was-saying Jodu that fact has-heard"
'Jodu has heard that Ram was talking to Shyam'

(9) [[[ram SEmer SOngge [$_{DET_i}$ je] kOtha bolchilo][$_{COMP_i}$ *e*]] jodu Se$_i$ kOtha Suneche]
"Ram Shyam-GEN with which word was-saying Jodu that word has-heard"
'Jodu has heard the thing(s) Ram was talking about with Shyam'

(10) *[[ram SEmer SOngge [$_{COMP_i}$ je][$_{DET_i}$ je] kOtha bolchilo] jodu Se$_i$ kOtha Suneche]
"Ram Shyam-GEN with that which word was-saying Jodu that fact has-heard"

The complement clause in (8) and the relative clause in (9) come out properly. Construal duly occurs in (9) and duly fails in (8), so that, in (8), the binder *Se* binds the COMP, while in (9),

the binder *Se* either binds the DET and overbinds the COMP or equivalently binds the COMP and subbinds the DET. However, in (10), there is a relative word *je* in the COMP slot, so that the DET *je* cannot get construed with COMP; what happens is that the binder *Se* binds COMP but cannot touch the unconstrued DET, which (with the rest of its NP) thus remains uninterpreted.

9.3 Direct alternative questions

Alternative questions as direct questions in Bangla and elsewhere fall into two main types – unary, as in (11-U), and multinary, as in (11-M).

(11-U) ram ki haSchilo?
"Ram whether was-laughing"
'Was Ram laughing?'

(11-M) ram ki haSchilo na kaMdchilo na cEMcacchilo?
"Ram whether was-laughing or was-crying or was-shouting"
'Was Ram laughing, or crying, or shouting?' (i.e. which of the three?)

Unary alternative questions like (11-U), which invite only a yes-no answer, are often called yes-no questions. Multinary alternative questions like (11-M), which in general do not invite a yes-no answer, must be distinguished from unary conjoined counterparts like (12), which do. The English glosses of (11-M) and (12) have different intonations (we use commas in the gloss of (11-M) to flag the difference). The Bangla sentences themselves differ in that the multinary-marking conjunction *na* of (11-M) contrasts with the disjunctive conjunction *ba* in (12).

(12) ram ki haSchilo ba kaMdchilo ba cEMcacchilo?
"Ram whether was-laughing or was-crying or was-shouting"
'Was Ram laughing or crying or shouting?'

(12), which has the surface constituent structure $[_{S'}[_{S}[_{NP}$ NP COMP$][_{V}$ V CONJ V CONJ V$]]]$, derives by COMP Preposing from (13), just as the somewhat marginal (14-ii) derives by COMP Preposing from (14-i) (its marginal status is due to the absence of the exact parallelism that coordinate structures require). Recall that COMP Preposing is optional and that therefore (14-i) is also a possible surface structure; it is well-formed as it meets the stringent criteria for parallelism imposed by coordinate structures.

(13) $[_{S'}[_{S}[_{NP}$ ram$][_{V}$ haSchilo ba kaMdchilo ba cEMcacchilo$]][_{COMP}$ ki$]]$

(14) i. $[_{S'}[_{S'}[_{S}[_{NP}$ ram$][_{V}$ haSchilo$]][_{COMP}$ ki$]][_{CONJ}$ ba$][_{S'}[_{S}[_{V}$ kaMdchilo$]][_{COMP}$ ki$]][_{CONJ}$ ba$]$
$[_{S'}[_{S}[_{V}$ cEMcacchilo$]][_{COMP}$ ki$]]]$
"Ram was laughing whether or was-crying whether or was-shouting whether"
ii. ??$[_{S'}[_{S'}[_{S}[_{NP}[_{NP}$ ram$][_{COMP}$ ki$]][_{V}$ haSchilo$]]][$ba$][[[$kaMdchilo$]][$ki$]][$ba$]
[[[$cEMcacchilo$]][ki]]]
"Ram whether was-laughing or was-crying whether or was-shouting whether"

'Was Ram laughing?, or crying?, or shouting?'

Neither (12) nor (14-ii) shows the effects of Conjunctionization, since the disjunctive conjunction *ba* is not a conjunction-complementizer, but just a conjunction. Recall that the obligatory rule of Conjunctionization can either move a +COMP +CONJ item into an empty +CONJ node or delete (in COMP position) a COMP-CONJ under identity (non-distinctness) with the contents of a non-empty CONJ position. Thus, for example, Conjunctionization deletes ki_i and ki_j (under identity with na_i and na_j respectively) in (15) to yield (16).

(15) ram haSchilo $[_{COMP}$ ki$][_{CONJ}$ na$_i$] kaMdchilo $[_{COMP}$ ki$_i][_{CONJ}$ na$_j$] cEMcacchilo $[_{COMP}$ ki$_j$]
"Ram was-laughing whether or was-crying whether or was-shouting whether"
'Was Ram laughing or crying or shouting?'

(16) *ram haSchilo na kaMdchilo na cEMcacchilo ki?
"Ram was-laughing or was-crying or was-shouting whether"

If we thus assume that the Conjunction *na* is non-distinct from the Complementizer-Conjunction *ki*, we not only explain the ill-formedness of (16) as a surface structure and the well-formedness of (14-i) as a surface structure, but also the grammaticality difference between (13) and its *na*-counterpart (17).

(13) ram haSchilo ba kaMdchilo ba cEMcacchilo ki?
"Ram was-laughing or was-crying or was-shouting whether"
'Was Ram laughing or crying or shouting?'

(17) *ram haSchilo na kaMdchilo na cEMcacchilo ki?
"Ram was-laughing or was-crying or was-shouting whether"

The reason for the ill-formedness of (17) is the same as the reason for the ill-formedness of the surface structure (15) – non-application of the obligatory transformation of Conjunctionization, which should delete the final *ki* under identity with the second occurrence of *na*. In order for our grammar to handle direct alternative questions properly, then, we simply need to claim that *na* and *ki* are, essentially, allomorphs of an interrogative Complementizer-Conjunction morpheme, and that *ba* is, in contrast, a Conjunction which is not a Complementizer.

We now need to justify the claim that *na* and *ki* are allomorphs. This justification will come from indirect alternative question data, data which also make it possible to describe the nature of the allomorphy of *ki* and *na*. One might wish to ask whether the Conjunction *ba* under consideration here is indeed distinct from the Complementizer *ba* which we have discussed before. I feel that, since the meanings of the Conjunction *ba* 'or' and the Complementizer *ba* 'on earth, at all, indeed' are quite different, and since the Complementizer *ba*, unlike the Conjunction, has special selectional restrictions discussed earlier, there is no need to marshal arguments in support of the claim that the two *ba*-s have nothing to do with each other. I will turn instead to indirect alternative questions in order to support and sharpen the claim about the relation of *na* to *ki*.

9.4 Indirect alternative questions

Consider an indirect alternative question such as the emphasized part of (18).

(18) **ram haSchilo ki na** ami jani na
"Ram was-laughing KI NA I know not"
'I don't know whether Ram was laughing'

It is possible to gloss *ki na* (also written *kina*) as "whether", and thus to claim that the indirect question in (18) is unary. It is also possible to insist on the two-word written form and gloss *ki na* as "or not", thus claiming that the indirect question is multinary (binary in this particular case). I shall refer to these as the Unary Option and the Multinary Option. I shall argue that one must take the Multinary Option. If one accepts my argument to this effect, then I will be able to produce evidence supporting the decision to treat *na* and *ki* as allomorphs of the same morpheme and clarifying the nature of their allomorphy.

Argument One: On the Unary Option, *ki na* counts as a Complementizer; in that case, it should undergo Complementizer Proposing and derive (19) from (18). But (19) is ill-formed. Therefore the Multinary Option must be taken.

(19) *ram ki na haSchilo ami jani na
"Ram KI NA was-laughing I know not"

Argument Two: On the Unary Option, *ki na* is an unanalyzable entity *kina*; in particular, it is not held to contain the negative element *na* (the one observed in *jani na* "know not"). Therefore, the Unary Option predicts that (20) should be well-formed, since it will not trigger the double negative filter. (Apologies for not expanding this comment; Bangla has such a filter; please take my word for it.)

This prediction also fails.

(20) *ram haSchilo na kina ami jani na
"Ram was-laughing not KINA I know not"
Intended reading: 'I don't know whether Ram wasn't laughing'

These arguments supporting the Multinary Option seem to me considerably more persuasive than the arguments which come to mind in support of the Unary Option. In that direction, one might argue, for example, that many users of Bangla write *kina* as a single word and that it therefore must be treated as an unanalyzable unit. This argument is weakened by the fact that many Bangla writers always write *na* "not" as part of the verb word, e.g. *kOrona* for our *kOro na* "(you-neutral-) do not". Although such writing together may be symptomatic of psychologically real word demarcation, it surely does not lead us to say that *kOrona*, for example, is a single morph! This reasoning carries over to *kina*. It is possible that the *na* is an enclitic on *ki* (or the *ki* a proclitic on *na*), so that *ki na* must be counted as a single word in some sense of the term 'word'. But that is no reason for regarding the unit *ki na*, if it is a unit, as unimorphic.

Another argument for the Unary Option might run as follows. One would argue that there is, in effect, a universal identity between 'whether X' and 'whether X or not', and that

therefore, even if Bangla insists on the presence of *ki na* "or not", such a language-particular wrinkle does not affect the conclusion supported by universal grammar, namely, that these indirect questions are unary alternative questions in the strict sense. Again, the general form of this argument seems to me suspect.

There are no cogent empirical arguments in favour of the Unary Option, as far as I can see. So, the Multinary Option stands.

One immediate positive consequence of the decision to adopt the Muitinary Option is that we have an explanation and not just a description of the ungrammaticality of (21). On the Unary Option, we would have had to handle (21) by means of lexical specifications. On the Muitinary Option, we allocate the task of starring (21) to the general principle that Unary Alternative Questions cannot be indirect.

(21) i. *ram haSchilo ki ami jani na
"Ram was-laughing whether I know not"
ii. *ram ki haSchilo ami jani na
"Ram whether was-laughing I know not"

In contrast, one apparently negative effect of this choice is that we now need to explain the obligatory absence of an interrogative complementizer in the question in sentences like (22) and (18); cf. (23) and (24).

(22) ram haSchilo ki kaMdchilo ami jani na
"Ram was-laughing or was-crying I know not"
'I don't know whether Ram was laughing or crying'

(18) ram haSchilo ki na ami jani na
"Ram was-laughing or not I know not"
'I don't know whether Ram was laughing (or not)'

(23) *ram haSchilo [$_{COMP}$ ki] [$_{CONJ}$ ki] kaMdchilo ami jani na
"Ram was-laughing whether or was-crying I know not"

(24) *ram haSchilo [$_{COMP}$ ki] [$_{CONJ}$ ki] na ami jani na
"Ram was-laughing whether or not I know not"

Only if a Conjunction precedes (not if it follows) a Complementizer non-distinct from it will Conjunctionization delete that Complementizer. So, in sentences like (23) and (24) where the relevant Conjunction follows the relevant Complementizer, we need a different process. I propose that the process be stated as follows.

(25) Alternative Conjunctionization (obligatory)
S.D.: [$_{\alpha}$X – [+COMP +CONJ +OPEN +VBL +INT] – Y – +CONJ – Z]
S.C.: 1 2 3 4 5 ➔1, 0, 3, 2, 5
(Condition: α is an indirect question)

One might think that this rule would incorrectly apply also to conjoined indirect constituent

questions, fatally removing the empty COMP node which chapter 8 has shown to be so vital to the proper semantic interpretation of constituent questions. Fortunately, (25) cannot affect constituent questions, since its action depends on there being an actual lexical item in COMP position, a lexical item which is marked +CONJ as well as +COMP. In a constituent question, the COMP position is empty, and therefore, not being +CONJ, fails to meet the S.D. for the second factor in (25), preventing application of (25).

One may object to the proviso 'α is an indirect question' and to the use of double bracketing in the S.D. I can only respond by saying that that's how the rule needs to be. Presumably it pays a high price to satisfy this need – in other words, the rule is presumably 'highly marked', peculiar to the language, etc. If one reflects that it is this rule which makes it necessary for all indirect alternative questions in Bangla to be multinary in form, unlike indirect alternative questions in many better-studied languages, then one ceases to find the conclusion that (25) is peculiar to the language distasteful. (25) certainly is closely linked with one of the most marked peculiarities of the language. The link between (25) and the Multinary Option is that, given the obligatory character of (25), it becomes possible to state the Multinary Option in an extremely natural fashion: one simply needs to add that (25) not only must apply to all conjoined indirect alternative questions (which have a CONJ in them) but to all indirect alternative questions. Thus, the condemnatory action of obligatory rules affecting structures which invite but block their application will now affect all unary indirect alternative questions.

Like the rule of Conjunctionization, which it supplements and mirrors, the rule of Alternative Conjunctionization either has the effect of moving the tangible contents of the COMP node into the CONJ node (which, by classical conventions of recoverability, must be empty for this to take place legitimately) or has the effect of deleting the contents of COMP under feature non-distinctness from the contents of CONJ in case both of these nodes are non-empty. For the notion of 'non-distinctness' see Chomsky (1965) and Bresnan (1976a). We need to notice that, thanks to the notion of non-distinctness, the Conjunction *na* and the Conjunction-Complementizer *ki* count as non-distinct for (27) to derive from (26) by Alternative Conjunctionization. The star on (26) indicates its status as a surface structure.

(26) *ram haSchilo ki na kaMdchilo ami jani na
"Ram was-laughing whether or was-crying I know not"

(27) ram haSchilo na kaMdchilo ami jani na
"Ram was-laughing or was-crying I know not"
'I don't know whether Ram was laughing, or, crying'

I take this to be a piece of evidence that *ki* and *na* are co-allomorphs of the same morpheme. Other considerations bearing on this alternation emerge from direct alternative question data.

(28) *ram haSchilo ki ki kaMdchilo?
"Ram was-laughing whether or …."

(29) *ram haSchilo ki kaMdchilo?
"Ram was-laughing or...."

(30) ram haSchilo ki, na kaMdchilo?

"Ram was-laughing whether or was-crying"
'Was Ram laughing, or, crying?'

(31) ram ki haSchilo na kaMdchilo?
"Ram whether was-laughing or was-crying"
'Was Ram laughing, or, crying?'

The contrast between the grammaticality of (30) and the ungrammaticality of (28) indicates that no process such as Alternative Conjunctionization is at work – that we are dealing, instead, with a process which discriminates between different allomorphs. The Isachenko-Aronoff haplology filter must be responsible. As in chapter 7, we notice again that the effect of this filter in Bangla is both to filter out the form with two identical affixes and not to let a form with one instead of two affixes take the place of the longer form. Thus, (29) is not allowed to do the work of (28), and, since there is nothing else (29) can do, it is rejected by the grammar. (This makes the situation somewhat different from the case discussed in chapter 7, where, although *kOkhon-o* "when-O" 'ever/sometimes', for example, was not allowed to do the work of the prohibited form *kOkhon-o-o* (which, if it had existed, would have meant 'even sometimes'), at least *kOkhon-o* had a right to exist, and to carry the interpretation 'sometimes/ever'.) I have added (31), which derives from (30) by COMP Preposing, for those speakers who find (30) relatively unacceptable (although I presume that there will be few, if any, speakers who fail to find (28) much worse than (30)); although the grammatical principles discussed in this volume have nothing to say about what might be responsible for the greater acceptability of (31) compared to (30), I feel that it is legitimate to appeal to (31)'s unquestioned acceptability to strengthen the case based on (30).

Returning to the question of *ki* and *na*, the above considerations indicate that these allomorphs are in partial complementation, as the structuralists used to say. In direct questions, only *na* occurs, since *ki* would cause the derivation to be rejected via the Isachenko-Aronoff filter. In conjoined indirect alternative questions where the second conjunct is the word *na*, only *ki* occurs (thus: in the sense 'I don't know whether Ram was laughing', *ram haSchilo ki na ami jani na* is okay, *ram haSchilo na na ami jani na* is ungrammatical), for the same reason. In these two sets of environments, then, *ki* and *na* are in complementation. But in a third set of environments (e.g. *ram haSchilo na kaMdchilo ami jani na*, *ram haSchilo ki kaMdchilo ami jani na* are both fine in the sense 'I don't know whether Ram was laughing, or, crying') *ki* and *na* are in free variation, there being no factors to constrain the occurrence of one or the other.

This is a relatively fortunate case, where we can derive the complementary distribution from general principles interacting with given properties of environments. This result is reminiscent of our earlier success in reducing the phenomenon of 'blocking' in the Bangla verb system to a collection of effects of various rules and configurations.

9.5 Yes-no questions and semantics

The above discussion to some extent clarifies both the fact that and the reasons why Bangla yes-no questions only occur as direct questions, where by 'yes-no question' I mean a question whose very format requires a yes or no answer. Thus, for example, in English, *Did he take it?* is a yes-no question, but the apparently synonymous *Did he take it or didn't he take it?* is not a

yes-no question, since it is couched in a format which would also accommodate questions not inviting a yes-no answer, like *Did he take it or did she take it?* This distinction between yes-no questions and other alternative questions would perhaps look slightly pedantic in a study of English. In Bangla, however, the distinction correlates with the syntactically crucial distinction between those alternative question (the yes-no ones) which contain an overt COMP and those (the rest) which, by Conjunctionization or by Alternative Conjunctionization, lose their COMP. Only yes-no questions, whose COMP survives long enough to be able to undergo COMP Preposing, can be aligned with the other types of clause whose syntax and semantics hinge on the COMP node.

So, continuing with our selective policy, we decide from this point on to regard Bangla yes-no questions as, for our purposes, privileged representatives of the class of Bangla alternative questions. Because of the way Bangla works, this happens to mean that we shall be excluding all indirect alternative questions. I hope the discussion above has sufficiently clarified the reason for this; it is an 'unintended consequence' of the decision to exclude clauses which do not clearly show what their Complementizer is doing.

As in the case of Declarative Pseudo-Complement (or Pseudo-Coordinate) Clauses discussed towards the end of section 9.1 (recall that we had not quite settled the terminological issue there), I shall exclude Interrogative Pseudo-Coordinate Clauses like *ram haSchilo na kaMdchilo* "Ram was-laughing or was-crying" (as in *ram haSchilo na kaMdchilo ami jani na* 'I don't know whether Ram was laughing, or, crying') from the scope of our discussion in this volume. The things which are pseudo about interrogative pseudo-coordinates are the same things which were pseudo about their declarative counterparts.

The semantic task, then, is to account for the construal and binding facts about sentences like *ram haSchilo ki?* "Ram was-laughing whether" 'Was Ram laughing?' Since we are only concerned with direct questions, for reasons explained above, the task is simple, and so is the apparatus which we bring to bear. In *ram haSchilo ki?*, we say that the COMP *ki* is a +VBL element which, carrying as it does the additional feature +INT, is bound by a semantic zero binder, as in constituent questions. The COMP node in a yes-no question, as in a complement clause, is filled by a word, and therefore cannot be construed with any interrogative determiners in the clause. Thus, *ke haSchilo ki?* "who? was-laughing whether" is ungrammatical (an English gloss that replicates the ungrammaticality would have to sound like

M M M H

'Was who laughing?'), as the only way its derivation could come out right would have been for *ke* "who?" to be construed with the filled COMP position.

This characterization of the semantics of yes-no questions makes an interesting prediction. Recall that an embedded constituent question can have a direct question effect if the question is an S object of the P *bole*, as in *ke haSchilo bole ramer dharona?* "who? was-laughing as Ram-GEN belief" 'Who does Ram think was laughing?' Now, it is pertinent to ask if the same thing can happen with an embedded yes-no question, to ask if one can say *SEm haSchilo ki bole ramer dharona?* "Shyam was-laughing whether as Ram-GEN understanding" with an intended reading which might be spelled out as 'Does Ram think A or B?', where A is the proposition that Shyam was laughing and B is the proposition that Shyam was not laughing. The answer, correctly predicted by our account, is: no, such strings are ill-formed. The reason is that P must take an S object – not an S′ object – and *SEm haSchilo ki*, containing as it does the COMP *ki*, is an S′.

One last point. There is the interrogative complementizer *bujhi*, which is not an Open

Word and therefore must count as –OPEN but +INT. Both the Open Interrogative complementizer *ki/na* and the non-Open interrogative complementizer *bujhi* must be absent from a COMP for that COMP to count as a potential target for a process of +INT construal, i.e. a COMP position in a constituent question. Thus, these complementizers are comparable to the complementizer *je*, which must be absent from relative clauses in all dialects of Bangla, not to the complementizers *naki* and *kina* which, in some dialects, may occur in the COMP position of a relative clause without preventing construal since *naki* and *kina* are not relative words. There are a few words like *re* and *go*, used in a very colloquial register and indicative of closeness between speaker and addressee, which may do for questions what *naki* and *kina* do for relatives – occur in COMP position without hampering construal and binding – except that it is not clear that *re* and *go* are complementizers.

9.6 The principal thesis and its scope

Turning from the elaboration and solution of individual sub-problems to the general problem, I would like to formulate the main contention of this dissertation as follows.

In Bangla, the characteristics of relative clauses (R) and of complement clauses (C) arise from two different but related 'procedures' (integral sets of rules) applied to the relative feature bundle J, the same 'procedures' which, if applied to the interrogative feature bundle K, yield respectively the characteristics of constituent questions (Q) and yes-no questions (Y). My use of the cover term Open Clauses for R, C, Q, and Y makes it useful to refer to this contention as the Open Clause Hypothesis or OCH.

In the formulation of OCH just given, the term Characteristics is to be understood as the combination of syntactic, morphological, and semantic properties. Elaboration of OCH must proceed both in a generalizing and in a particularizing direction.

To begin arbitrarily with the latter direction, the first of the 'procedures' – the one which, applied to J, gives the characteristics of R, and, applied to K, gives those of Q – is as follows. Insert a lexical entry with the feature bundle J (or K) into the DET position of an NP or AP. Subject the subtree thus obtained to regular processes of the grammar in the different directions defined by general linguistic theory. In the direction of morphology spell out the appropriate J-word (or K-Word) DET (with due regard to any zero Noun which may be adjacent) as a specific morph. In the direction of transformational syntax, subject the DET to Phoric Postposing, optionally. In the direction of semantics, interpret the feature bundle J (which is, to be precise, +OPEN +VARIABLE –INTERROGATIVE) or K (which is +OPEN +VBL +INT) as required: first, construe it with the COMP node in the minimal S′ dominating it so that the construed DET becomes a dependent variable and the COMP an independent variable; then, subject it to direct binding or subbinding, and, if necessary, transfer its boundness to COMP via the overbinding convention – where, in all cases, it is understood that sentence grammar does not identify the binder of any J element but does identify (as a semantic zero) the binder of every K element.

The procedure just described derives its unity from the unity of the features OPEN, VBL, INT, which are responsible for the goings-on in all sectors of sentence grammar as far as Open Clauses are concerned.

The second procedure – the one which, applied to J, produces C characteristics, and, applied to K, produces characteristics of Y – is as follows. Insert a lexical entry with the feature

bundle J (or K) into the COMP position of an S′. Subject the tree or subtree thus built to regular processes of different sectors of the grammar. In the direction of morphology and phonology, there are no decisions to make, since the J-Word and the +INT Words which can occur in COMP position are all mono-allomorphic. In particular, the J-word is *je* and the most commonly used +INT Word (the one which is a K-Word and which also shares the +CONJ feature with the C-type COMP *je*) is *ki*. In the direction of transformational syntax, subject the COMP to COMP Preposing, optionally, and, if appropriate, to Conjunctionization or Alternative Conjunctionization, in which case we will not derive a C or a Y. In the direction of semantics, interpret the bundle J or K as required: subject it to direct binding – it being understood that sentence grammar does not identify the binder of the J element, and does identify the binder of the K element as a semantic zero.

Turning now to the task of elaborating OCH in a generalizing direction, I point out that both DET and COMP, the crucial nodes involved in the above characterizations of the procedures, are Specifier nodes in the sense of 'specifier' defined in the bar-notational literature. Given the fact that, in Bangla, COMP occurs at the end of its S′ while DET occurs at the beginning of its phrase, it becomes possible to see both COMP Preposing and Phoric DET Postposing as special cases of the process of 'moving the SPEC around within its mother phrase'. Both COMP and DET are cross-referentially responsible for their mother phrase (respectively, for S′ and for NP or AP) in the sense that the values on features like VBL and INT in the Specifier determine the cross-referential status of the entire phrase. An S′ with a +INT COMP, for example, is a question, and an NP with a [+VBL –INT] DET is a relative phrase.

The crucial points of the elaboration as it has proceeded so far can be epitomized as follows.

(32) The Open Clause Quadrangle

	DET-Open Clause Types	COMP-Open Clause Types
+INT	Constituent Questions	Yes-No Questions
–INT	Relative Clauses	Complement Clauses

Diagram (32) lists the conventional names of the clause types obtained when one places +OPEN +VBL words with, respectively, +INT or –INT as an additional feature, in, respectively, DET and COMP position.

Given the above general picture we can distinguish the Specifiers of NP, AP, and S′, which participate in Open Clause affairs, from the Specifiers of PP (do they exist?) and VP (Tense, Aspect, Negation?), which do not. This distinction is drawn in section 8.10, where the term 'hypophoric' is suggested for this property of the categorial types N, A, and S, and where a link with child language acquisition is pointed out.

We can also ask what the property shared by all hypophoric nodes is which lies at the heart of their manner of linking up with one another syntactically (through embedding and coordination) and semantically (through construal and binding). The answer which the present study suggests to this question is that the crucial property is the ability of all and only hypophoric nodes (or rather of their specifiers) to take a plus value for the feature VBL and thus to make it possible and necessary for other hypophoric nodes to be construed with and (in

the case of relative clauses and complement clauses, that is, in the case of the J element) to bind the variable node or (in the case of questions, that is, of the K element) to be associated with a semantic zero which binds the variable node. It seems that the syntactic embedding relation which holds between, say, a relative or complement clause and its matrix provides a frame which facilitates the processes of construal and binding but does not dictate their course. Thus, it has been possible to write a viable grammar of the Bangla open clause system without positing any direct link between the syntax and the semantics of –INT clauses. The non-linkage is perhaps even clearer in the case of questions, +INT clauses, where the zero binder at the heart of the semantic analysis does not even exist on the plane of syntax, if one accepts my second and not my first approximation to the true rule of Interrogative Binding.

Thus elaborated, the principal thesis of this volume, OCH, obviously straddles the fence (if any) between general linguistics and the grammar of Bangla, since the notions SPECIFIER, VARIABLE, etc. are part of linguistic theory and pertain to all languages. In section 9.7 I will collect some loose ends left in this inquiry as it has progressed so far and indicate possible ways to follow up these leads. But first I must characterize the scope of OCH in positive terms, by giving examples of phenomena and principles which OCH is directly pertinent to. This characterization, again, must be given both in general and in particular terms. I will begin with the general aspect.

Let me outline the context first. Considerable attention has been paid in the recent literature to precisely the clause types I am concerned with. The discussion has proceeded within the framework set by the controversy between Noam Chomsky and Joan Bresnan over the nature of language and the appropriate goals and methods for linguistic research. Chapter 10 will take up the 'Chomsky-Bresnan debate' in some detail. Here it may suffice to note that parties to this debate have concentrated on the putative gap in a 'wh' clause, the gap which Chomsky sees as an extraction site (a trace of movement) and which Bresnan, at least in the case of English free relatives, sees as a deletion site (a trace of deletion under identity over a variable). For languages which we may call Open-Topical languages – since they place their Open ('wh') phrases typically in a Topic ('preposed') position – this massive research effort is bearing direct fruit. But it remains unclear what bearing, if any, the results of this research have on languages, which we might call Open-Atopical languages, where an Open phrase, instead of occurring typically in a Topic position, occurs where its non-Open congeners occur – in any position, at the beginning, middle, or end of a clause.

In this context, the Open Clause Hypothesis serves to put the current mainstream research effort 'in its place' with respect to Open-Topical and Open-Atopical languages. Since OCH applies to both classes of languages, it is conveniently universal (unlike the thrust of the mainstream research, which for the most part is relevant only to Open-Topical languages). But OCH also incorporates at its very core the concepts of a Variable node (to be bound) and of linkage (called Construal) between the Independent Variable COMP of a clause and its one or more Dependent Variable DETs. These concepts allow us, on the basis of OCH, to glimpse a deeper explanation of the regularities about Open-Topical languages noted ad nauseam in mainstream research.

The latter research dwells on the link between, for example, the wh-phrase in the COMP position of an English constituent question and the open phrasal site (the site of deletion or extraction associated with this phrase). The Open Clause Hypothesis makes it apparent that the relation (seen as one of construal) between the COMP position and some other phrasal position which it c-commands obtains both in an Open-Atopical language like Bangla, where it

is the COMP position that must be empty and the open phrasal position that must be filled, and in an Open-Topical language like English, where it is the other way round. Thus, the Open Clause Hypothesis helps us understand, in terms of a general feature of language (a 'language universal'), what mainstream research merely notes on the level of description of a property of many but not all languages. Instead of simply noting that, in many but not all languages, an Open phrase in COMP position must be related to its open phrasal site within the domain of this position, we now see this fact as a special case of the general need to associate an Open COMP position with a clause-internal Open phrasal position for the work of a DET-Open rather than COMP-Open (a 'phrasally' open rather than 'clausally' open) clause, the work of a constituent question or relative clause, to proceed properly.

Shifting now to a particularist gear, the most direct entailment of OCH is the parallelism claim – the claim that throughout the world, in more cases than could arise from coincidence, we will see a formal and functional parallelism between relative clauses and constituent questions, between relative and complement clauses, etc. Obviously, this claim will need to be tested over the length and breadth of languages. I can only give a few examples to show that it will not be a wild goose chase.

A particularly striking example of this sort is the use in Homeric Greek of a form identical to the neuter singular nominative/accusative relative pronoun *hó* in the role of introducer of complement clauses. Saul Levine has found for me an example where *hó* alone is so used – //gignọ̄́skon *hó* hoi autòs hupeírek^{h}e k^{h}eîras Apóllọ̄́n// 'realizing *that* Apollo himself was holding his hands over him' (*Iliad* 5.433) — and three examples (two of them Homeric and one post-Homeric) where the complementizer *hó* appears fused with the indefinite pronoun *ti*, also neuter singular nominative/accusative: //eíp^{h}' *hóti* hoi sộs eimi kai ek Púlou eilẹ̄́lout^{h}a// 'say that I am safe and have come from Pylos' (*Odyssey* 16.131); //ẹ̄́ggeil' *hótti* r^{h}á hoi pósis éktot^{h}i mímne pulā́ọ̄n// 'he reported *that* her husband was staying outside the gates' (*Iliad* 22.439); //apokrīnámenoi *hóti* pémpsousin họ̄s autoùs présbeis// 'replying *that* they would send ambassadors to them' (Thucydides 1.90.3). The//...// notation is an ad hoc marker of transliteration.

A better-known example comes from Modern German, where a form identical to the neuter singular nominative/accusative pronoun *das* plays the same role. There are analogous phenomena in plenty of other well-studied western languages, as they stand and in their older forms as well.

One more instance of formal kinship between relative and complement clauses is observed in French sentences such as *Je le vois qui danse* "I him see who dances' 'I see him dancing', where the most natural reading of the sentence is one in which the object of 'see' is 'him dancing', a proposition, rather than 'him', a term with the property of 'dancing', leading us to speak of a complemental semantics not normally associated with the relative clause syntax observed in *qui danse*. Full-scale generative analysis of this French 'pseudo-relative' construction (as it has been called by Radford (1975)) has yet to be carried out. Considerations based on it may eventually lead to a modified version of OCH interrelating the French data and patterns now covered by OCH. The French COMP *que*, the normal marker of complement clauses, coincides with a form of the French relative pronoun *qui*. On French pseudo-relatives, see also Mira Rothenberg (1972). I am indebted to Paul Modini (p.c.) for these references.

We find a Japanese analogue to the French pattern. A detailed description of the Japanese pattern, mentioned to me by Kuroda in 1979, appears in Kuroda (1977). One example is *Taroo wa ringo ga sara no ue ni aru no o mita* "Taro WA apple GA plate NO on NI be NO

O saw", where, if the NO before O is interpreted as a complementizer, the string means 'Taro saw that an apple was on the plate', but, if that *no* is instead seen as indicating a 'pivot-independent relative clause', the string means 'Taro saw an apple which was on the plate'. See Kuroda (1974, 1976) for the rest of his study of pivot-independent relatives and for references to analyses of other languages exhibiting similar phenomena. It is interesting that certain occurrences of *no* in Japanese seem indeterminate as to whether they instantiate a Complementizer or a Pronoun. Cf. (8)-(9) above for a similar indeterminacy in Bangla.

These facts from Greek, German, French, and Japanese all exemplify isomorphism between R and C. R also parallels Q. In English, French, and many other languages, the system of relative words which figures in R and the system of interrogative words used in Q are more or less the same system – 'wh' words in English, 'qu' words in French, etc. (the qualification 'more or less' is an allusion to specialized words like the exclusively interrogative English *how*, whose relative counterpart *as* also doubles as a demonstrative, and the exclusively relative French *dont*). I will use the term 'Open Words' as a neutral cover term for this system. In many of these languages, the set of Open Words is subject to a rule which, in classical transformational grammar, is thought to move the minimal phrase containing an Open Word to a clause-initial position, and, in some of the work by Bresnan and other strict lexicalists, is thought to delete material within the clause under identity with a clause-initial Open Word.

Turning now to the parallelism between constituent questions and yes-no questions, we need not look far. Questions are widely regarded as a natural class. In English they trigger subject-auxiliary inversion in root sentences. In French they trigger right-adjunction of a pronoun subject or clitic to the first finite verb of the clause. Q and Y also resemble each other in semantics (they both 'ask') and phonology (they both show some sort of 'question intonation').

Finally, in well-studied languages, C and Y share several features. In English, both C and Y take part in a process which we may continue to call 'Extraposition' on historical grounds, regardless of the real nature of the process ('intraposition', 'true' extraposition, interpretation in place, or whatever).

Even such a perfunctory inspection of the range of relevant data, I hope, gives a sense of how much there is to investigate when testing the validity and refining the formulation of the Open Clause Hypothesis on the level of language universals.

I will now draw attention to areas where follow-up research is needed in that the present study leaves work unfinished or suggests new questions.

9.7 Suggestions for follow-up

The clauses which I have variously called Pseudo-Complement or Pseudo-Coordinate Clauses obviously deserve further study. The heuristic of the present study has been to search for J and K elements in the class of Complementizers (basing the inquiry on Determiners, the 'home ground' of J and K); this heuristic directly suggests that we extend the chase into the territory of Conjunctions, taking advantage of the fact that *je* and *ki* are simultaneously Conjunctions and Complementizers. Such further study would have to base itself on a good analysis of the coordination system.

Another obvious loose end is Right Relative constructions, which I have neglected because I have found it difficult to state and substantiate anything significant about them that

has any continuity with Left Relatives. On the one hand, they are similar enough to Left Relatives that what I have said about the latter carries over to them at the levels of semantics and morphology, as far as I can see, though less so in syntax. On the other hand, Right Relatives, being not as strikingly parallel to constituent questions as Left Relatives are, lend themselves less well to the logic of the present inquiry. I feel – perhaps mistakenly – that it will be possible to extend the results obtained for Left Relatives to Right Relatives. I anticipate a need to use Mark Baltin's (1978) construal schema for postposed modifying structures when carrying out this extension of the inquiry.

The two tasks mentioned are simply tidying up tasks engendered by the way the present inquiry was conducted, and may prove to be theoretically unrewarding. There is a third task, which also directly follows from the incompleteness of the present work, but which goes beyond such tidying up. This is the task of explicating, in Bangla semantics, the relation which my dependent variables marked by J/K determiners bear to universal and existential quantification. In this connection it will be necessary to look closely at sentences like *ami tomaY jOto Suniyechilam gan tar bOdole ami cay ni kono dan* "I you-OBJ as-many had-sung songs they-GEN in-exchange I wanted not any return" 'I sung all the songs I sung for you not in the hope of any concrete return' where the relative word (in this case *jOto* 'as many') in effect means 'all'.[21] Alongside these, one will have to look at Indefinite words and Indefinite uses of Interrogative words, extending my study of this matter. It seems plausible to expect that the universal and existential quantifiers will turn out to be semantically derivative constructs which build, respectively, on relative and interrogative construal schemata. This result, when firmly obtained, will radically alter the relation between linguistic and logical research.

One piece of Japanese evidence to approximately this effect (i.e. to the effect that the quantifiers are derivative constructs) was provided by S.-Y. Kuroda in 1979 when he pointed out to me that the element *ka*, which is in Japanese what *o* is in Bangla – an element which forms indefinite words from an interrogative stem – is identical to the interrogative conjunction-complementizer *ka* which, in its conjunction use, has the disjunctive meaning 'or'. Kuroda went on to say that, in the structure of classical predicate logic, the disjunctive 'or' element corresponds to the existential and the conjunctive 'and' element to the universal quantifier, in that, given a set with members X, Y, Z, the expression 'all members' means 'X and Y and Z' and the expression 'some member' means 'X or Y or Z'. The fact that this disjunctive-existential *ka* element in Japanese is also an interrogative complementizer translatable as 'whether' seemed to Kuroda, as it does to me, to support the idea of some special link between the semantics of interrogation and of existential quantification, perhaps via disjunctive coordination.

An argument for the idea of a special tie between the semantics of relatives and of universal quantification via conjunctive coordination can be constructed, perhaps, on the basis of the widespread, and traditional, idea that at least some relative clauses (certainly 'non-restrictive' or 'appositive' relatives, and perhaps some others) are to be interpreted as coordinate conjoined propositions.

---

[21] A 2020 addition: A relative word plus the emphasizer /i/ often carries the denotation 'wh...ever' (/jai/ 'whatever', /jei/ 'whoever', /jekhanei/ 'wherever', /jOkhoni/ 'whenever'), approaching universal quantification. Recent work on these matters, in several languages, was optimally consolidated by Rahul Balusu (only a few years before his tragic passing).

To move closer to concrete facts (rather than classes of facts), one can argue for the same conclusion on the basis of the dual role of words like English *both* which firmly underpin the hint that conjunctive coordination and universal quantification are closely akin. *Both*, as discussed in a quite different connection earlier in this volume, is a close kin of *all* and *every*. It is clearly a quantifier in phrases like *both sailors* or *both of the sailors*. On the other hand, as discussed in Dougherty's definitive work on coordination in English (1970, 1971), *both* is also part of a coordinating conjunctional pair *both...and...*, as in *John both wrote and mailed the letter*, an inappropriate context for a quantifier (**John all wrote and folded and mailed the letter*). If this were a quirk peculiar to *both*, we might have pleaded homonymy. But the phenomenon is systematic. *Either* and *neither*, like *both*, occur as quantifiers: *Either of the spouses may sign*, *Neither of the spouses knows*. Again, these words also occur as part of a coordinating conjunctional pair — *either...or...*, *neither...nor...*: *John either sent or got a letter*, *John neither sent nor got a letter*. In semantic inquiry, one must design one's 'semantic alphabet' in such a way that the kinship between coordination and quantification (between conjoining and universal quantification and between disjoining and existential quantification) enters into the constitution of the notions 'coordinator' and 'quantifier'. No current theory of semantics meets this requirement.

Another argument in favour of the idea of linking the semantics of relatives to universal quantification, this time via definiteness, arises in Bangla. It is known among Bangla-English translators that the J element has a definitizing force; Tagore (1938) mentions it in his notes on the Bangla language. Thus, while English allows you to say *The chair in the garden is sturdy*, we cannot say *bagane ceYarTa mojbut* "garden-LOC chair-item sturdy" (grammatical on the odd and irrelevant reading 'The chair is sturdy in the garden (but, presumably, not elsewhere)'), but must instead choose between *baganer ceYarTa mojbut* "garden-GEN chair-item sturdy" 'The chair in the garden is sturdy' and the somewhat bizarre *bagane je ceYar SeTa mojbut* "garden-LOC which chair that-item sturdy" 'The chair which (is) in the garden (is) sturdy'. Chomsky (1977a) has suggested that the notion of definiteness (at least when adjacent to a restrictive relative clause) is to be analyzed in terms of universal quantification – witness the 'universal' force of *the books we ordered* ('*all* the books we ordered'). Anyone seeking to relate universal quantification to the semantic properties of relatives in Bangla will do well to accept and add precision to Chomsky's idea.

So far, I have been listing isolated arguments in support of the relative half and the interrogative half of the idea that it is the interpretation of dependent variables that underlies quantification (both universal and existential). I now turn to a theoretical consideration which begins to bridge the gap between the two halves.

The essential idea is that under appropriate conditions (intonation, context, etc.) a variable located in a phrasal rather than sentential SPEC (i.e. in DET and not in COMP), if it is bound by a particular antecedent phrase present in the syntactic structure, gets a universally quantified reading, but, if it is bound instead by nothing syntactically present but only a semantic zero (and thus, in effect, is not bound), gets an existentially quantified reading.

Stated in this form, the idea sounds remarkably like a description of trace and PRO_{arb} in Chomsky (1978). Trace, which is bound by a specific present antecedent, normally has the reading 'the ... N', while PRO_{arb} , which is 'arbitrarily bound' (i.e. bound by nothing syntactically present, as in *It is unclear who* PRO_{arb} *to visit t*), has the reading 'some ... N'. In the form of 'zero produced by deletion under identity' or 'base-generated zero', trace has its counterparts in alternative systems, and PRO_{arb} may carry over directly into their analyses of

English. In other words, what I am referring to is not an exclusive feature of Chomsky's work alone.

The general principle involved in these cases seems to be that an anaphor which is 'concretely bound' (i.e. bound by something in the structure) is universal, while an anaphor that is 'abstractly bound' (i.e. bound by nothing in the structure) is existential. To the extent that this principle can be made precise and firmly founded, it will become possible to derive the 'logical primitives' of quantification from anaphoric foundations. Consequently, linguists' uncritical acceptance of the hierarchy of constructs as given in standard logic will be undermined. It will be seen that linguistic research can illuminate the empirical structure of logical vocabulary as well as the other way round.

One may object that my analysis of Bangla indefinite expressions as involving binding of the K element by an *o* quantifier present in the structure controverts the idea that 'binding by something present in the structure' yields universal quantification, since indefinite forms are manifestly existential rather than universal. But that is exactly what I mean by saying that the idea needs to be made precise. In this case I would make it precise by noting, for example, that the binding of *o* by K is part of the lexical semantic content of the set of indefinite words, not part of the set of interpretive semantic rules which process syntactic structure, and that therefore the formulation of what yields universal quantification should be altered so that 'present in the structure' refers only to syntactic structure, not to intra-word lexical structure.

Turning from the third task arising from the present inquiry to the fourth task, let us consider another thread that has to do with definiteness and relatives, using our old friend *the chair in the garden is sturdy* as a peg to hang the discussion on. This time we look at the genitive-laden translation *baganer ceYarTa mojbut* "garden-GEN chair-item sturdy", flagging the fact that the use of the genitive "garden-GEN" has the effect of definitizing the word "chair", and that this is equivalent to the definitizing effect of a relative construction. Add to this the fact that, corresponding to the English relative infinitival construction as in *the knife with which to cut diamonds*, there are two Bangla constructions, one with a relative (but a finite) clause and the other with a non-finite (gerund) form of the verb but with a genitive rather than a relative construction: (i) *je churi diye hire kaTte hOY Sey churiTa* "which knife with diamond to-cut is that knife-item", (ii) *hire kaTar churiTa* "diamond cutting-GEN knife-item". Again, a genitive and a relative function equivalently in the role of definitizer. This equivalence is unsurprising in the context of Indian grammatical terminology. In the Sanskritic tradition, the term *sambandha-pada*, used for a word in the genitive, means literally 'relation-word'; what could be more natural than an equivalence between Relation Words and Relative Clauses?

The definitizing role of genitives is known from other languages, of course, e.g. English (*my brother* is like *the brother* in many respects). Joseph Malone (1977) proposes to regard definite noun phrases as in some sense self-possessed. What is not known is the nature of the relation between the relative as definitizer and the genitive as definitizer. That there is a relation to be studied here is indicated by a Japanese datum, the fact that *no* (which, we have seen, marks a relative construction) is also the normal genitive Case marker. Kuroda (1974) goes on to note that it is well known in Japanese linguistics 'that the conjunctions *ga*, *wo*, and *ni* were derived from the case markers *ga*, *wo*, and *ni*', suggesting that there is a wide range of connections to be explored between the grammar of Case and that of CONJ-COMP elements even in languages which, not being Japanese, do not wear this set of connections on their sleeve.

In fact, Japanese is not the only language to display a genitive-relative connection.

Dixon (1966) reports that Dyirbal and Gumbainggar are languages where 'possessive phrases can be dealt with as a special case of the relative clause construction' (1966:35).

A fifth task, for Bangla and other languages, is to study borderline cases of non-interrogative open clauses. For example, the Bangla word *jate*, morphologically a locative of *ja* 'what' and used in that capacity in sentences like *je jate anondo paY Se ta-i korbe* "who what-LOC joy gets (s)he that-EMP will-do" 'People will do whatever they take pleasure in', has a special use associated with the meaning 'so that'; in that capacity, it serves as a non-interrogative open complementizer: *otithira jate nijer nijer icche mOto khabar beche nite pare Sey jonne e rOkom bEbostha korechi* "guests so-that own own wish according food choose-and take can that for this sort arrangement have-made" 'I have made this sort of arrangement so that the guests can pick whatever food they want'. Now, consider an intermediate case like *apnar **jate** Subidhe hOY apni ta-i korun* "you-GEN **JATE** convenience is you that-EMP do"; the string is ambiguous between 'Do whatever is convenient for you', taking *jate* to mean "what-LOC", or as 'Do (something) so that it is convenient for you', dropping an argument (a standard possibility in Bangla) and taking *jate* to mean "so-that". Expository convenience leads me to describe this string as ambiguous, with two distinct meanings; others may prefer to regard it as a vague sentence, with a continuum of meaning definable in terms of two poles. In either case, the exact relation between the roles of *jate* is an unsolved problem.

One finds slightly different unsolved problem in the pair *jehetu ... Sehetu ...* The noun *hetu* 'reason, cause' no longer survives on its own in normal Bangla (it still exists as a technical term in philosophy), but its derivative *Ohetuk* 'pointless' remains, making it moderately reasonable to regard *jehetu* and *Sehetu* as even synchronically DET N sequences with an understood (presumably Locative) Case suffix. (However, interrogative **kihetu/konhetu*, proximal **ehetu*, and distal **ohetu* do not exist.) Given the analogous form *je karone* "which reason-LOC" and *Se karone* "that reason-LOC", which occur in sentences like *ager bar je karone juddho themechilo e bar Se karone thambe na* "last time which reason-LOC war stopped this time that reason-LOC will-stop not" 'The war will not, this time, stop for the same reason as the reason for which it stopped last time', one would expect a similar reading for *ager bar **jehetu** juddho themechilo e bar **Sehetu** thambe na*. Instead, this string yields only the reading '**Because** the war stopped last time, **therefore** it won't stop this time', with *jehetu* and *Sehetu* read as 'because' and 'therefore' respectively. I find this surprising and would like to find an explanation. [A 2020 addendum: the question remains unanswered.]

In the above cases, we have a prima facie relative NP functioning as a complementizer. Like other non-interrogative open complementizers, *jate* and *jehetu* are intolerant of relative phrases in the clause: **jara jate icche mOto khabar beche nite pare* "who so-that wish according food choose-and take can", **jehetu je juddho themechilo* "since which war stopped". This functional switch (from relative NP to COMP) must form part of the explanation of the grammar of *jate* and *jehetu*.

It is even harder to claim that *jodi* 'if' and *jodio* 'though' are relative words; there is insufficient warrant in Bangla synchronic grammar for an analysis in terms of DET N. Their Sanskrit etyma *yadi* 'if' and *yady-api* 'though' had a sufficiently obvious relatedness with the word *yadaa* 'when' (cf. the two meanings 'when' and 'if' of German *wenn*) which fell into the series *tadaa*(*-niim*) 'then', *idaa-niim* 'now', *kadaa* 'when?', *kadaa-'pi* 'ever', *kadaa-cit* 'sometimes', etc. In Bangla, the sequent that follows up after *jodi* 'if' is *tObe* 'then' (we also get *tahole* 'then', which freezes the phrase *ta hole* "that happening" into a word), while the sequent that corresponds to *jodio* 'though' is *tobu* 'still'. The isolated interrogative word *kObe*

'which day?/when' and the archaic relative word *jObe* 'when' give some paradigmatic support to an analysis of *tObe* and *tobu* as NPs, and, by implication, to a similar analysis of *jodi(o)*. One may add, with no psychological reality claims, the alternation of *kObe* 'when?' and the archaic *kobhu* 'ever' as evidence for a morphophonemic process which turns *kObe-o* and *tObe-o* into the parallel *kobh-u* and *tob-u*; recall that the indefinitizing *o* when attached to *ke* 'who?' and *ki* 'what?' yielded the u/w-laden output *kew* 'someone' and *kich-u* 'something', showing that the high *u* allomorph for *o* is independently available. If *tobu* 'still', the sequent companion of *jodio* 'though', is derived from or intimately cognate to *tObe-o* "then-even" 'even then', then one begins to feel that *jodi* and *jodi-o* are somehow akin to the *kObe-tObe* series which is no longer synchronically a true series at the morphological level, having lost several members (it has lost the relative *jObe* and the proximal *Ebe*). This does not suffice to establish *jodi* as an NP. But it does, to my mind, warrant the inclusion of *jodi* in the class of J-Words, although this move raises a problem, in that *jodi*, with its derivative *jodio*, thus becomes the first J-Word which can take an *o* suffix of any sort – jeopardizing the basis on which the class of Open Words was established on the morphological level. Perhaps this problem arises only on the incorrect assumption that *jodio* is synchronically bimorphic. There is evidence that *tobu* is not synchronically *tObe-o*: if it were, the word *tobu-o* would be impossible (by the Isachenko-Aronoff haplology principle: **tObe-o-o*); but *tobu-o* does occur, as a stronger variant of *tobu*; if we must therefore conclude that *tobu* is synchronically unimorphic, then so, perhaps is *jodio*.

If we set aside these morphological concerns, my main problem with clauses containing *jodi* 'if' and *jodio* 'though' is that, if *jodi* and *jodio* are J-Words, then clauses containing them can only be Complement Clauses, since these J-Words are syntactically conjunction-complementizers and not phrases or phrasal specifiers. Why is this a problem? For several reasons. First, the complementizer *ba* occurs with *jodi* (*jodi ba* "if BA" 'if by any chance'); if we were to recognize *jodi* as a complementizer, then this would be the only violation of the principle that a clause should not have two complementizers. Second, there would then be a need, given the profusion of J-Word complementizers, to analyze the similarities and differences between complement clauses containing one or another of the complementizers *je*, *jodi*, *jodio*, *jate*, and *jehetu*. It may prove impossible to carry out this task without destroying the unity of the notion 'complement clause'. In particular, *jodi*-clauses exhibit marked peculiarities in the areas of Tense and Negation, although *jodio*-clauses don't. (The peculiarities with respect to Negation are easy to explain in terms of there being a two-word item *jodi na* 'if not' which launches its *na* into a clause by a floating rule. The tense facts are tougher to handle.) Third, this prolixity in the class of Complement Clauses would match nothing in the unenlarged class of Yes-No Questions, raising the question 'Why the discrepancy?' Fourth, with *jate* and *jehetu* there is at least a tenuous contact with plausible NP sources for the complementizers. But with *jodi* the link with the NP system is synchronically non-existent. I had to resort to etymology to show that there had once been a link. Thus, the semantics of *jodi*- or *jodio*-clauses has no anchor in that of specific relative clause types. Fifth, different types of complement clauses[22] would line up with different areas of the verb system

[22]A 2020 comment: it was an error in my thesis not to distinguish the notion of 'clause built around the complementizer /je/' sharply from the notion of **'complement** clause'. Many instances of the former are **adjunct** clauses and should have been studied separately. In that period, many authors still maintained that *that you can't go there on foot*

for no apparent grammatical reason – *je*-clauses with gerunds, *jodi*-clauses with verb phrases (not studied in this volume) where the verb is in the protatic *-(y)le* form, *jodio*-clauses with verb phrases in *-(y)le-o*, and *jate*-clauses with verb phrases in *-(y)te*. By 'lining up with' I here mean a rough but systematic paraphrase relationship, which, even if it has no syntactic significance, must be accounted for in Bangla semantics.

All I know is that these problems will have to be tackled. I have no suggestions to offer as to how to go about doing so. A sixth task, interwoven with some of the others, is to relate the sort of relative clauses studied in this volume (involving the binding of one or more Dependent Variables in the clause) to what Kuroda, in the references cited in section 9.6, calls pivot-independent relative clauses, which seem to have no Dependent Variables in my sense.

A seventh task is to study the sequent complementizer-conjunction *to* in Bangla in its relation to other members of the sequent system, which deserves to be studied as a system. I have not undertaken that study here because the properties of the sequent system interlock sentence grammar with discourse grammar in ways which we can profitably study only if we have an overall idea of discourse grammar.

It is to be expected that research in the areas mentioned, when carried out, will revise the details of the analyses I have proposed without, I hope, shaking their foundations. For reasons of space, I have not outlined possibilities of follow-up that involve psycholinguistics and sociolinguistics.

9.8 On the way to the end

The suggestions just made, going off as they do in many directions, call for a variety of investigations by people with overlapping expertise. Instead of directly attempting to launch such a massive effort, chapter 10 will less ambitiously address itself to part of the set of questions popularly known as the Chomsky-Bresnan debate, because some (in particular, Chomskyan) answers to these questions are at odds with a few of my proposals about Bangla, and because independent arguments favouring Bresnan-type alternatives to Chomsky's work lend indirect support to the orientation within which the present inquiry has been conducted. However, one by-product of chapter 10 will indeed be a suggestion pertaining to the fourth of the tasks in section 9.7 – the task of relating genitive definiteness to relative definiteness.

Chapter 9 will thus be the last chapter to concentrate on specifically Bangla linguistics. Let me, then, before ending the chapter, outline the starting points and basic findings of this study. The account which follows is slightly different in emphasis from the one given in chapter 1, which it may be useful to re-read at this point.

The inquiry which ends here started out on the basis of the observation by Rabindranath Tagore (1938: 98, 116) that the relative pronoun *je* and the interrogative pronoun *ki* also have non-pronominal uses in which they underpin the complement clause construction and the yes-no question construction, respectively. The first half of this observation was independently made by Duncan Forbes (1861: 154), who noted that 'the relative *je* is used as a conjunction, "that", in a manner similar to the use of *huti* in the Greek "New Testament".' I set out to find a

---

was the (extraposed) complement of *so* in *The village is so far away that you can't go there on foot*; the complement-adjunct distinction was less clear and robust than it is today; the failure was not mine alone. But it **was** a failure.

way to place the generalization noted by Tagore in a contemporary theory of grammar. Kuroda's (1968) pioneering work on the grammar of relative and interrogative pronouns helped me to formulate an account. Kuroda has shown that the relative/interrogative character of the pronominal NP resides in its determiner and that, at least in English, on all grammatical levels (morphology, syntax, semantics), the relative element is akin to the interrogative element in a way which can be expressed by positing an archi-element WH shared by relatives and interrogatives. It turns out that Kuroda's finding carries over even to a language like Bangla, where relatives and interrogatives seem to have morphologies quite different from each other, an interesting and surprising fact. My analysis expresses this fact by using a +OPEN feature both for relative and for interrogative determiners. Another trait of Kuroda's analysis which serves as a working hypothesis in my research, and seems to have entered the store of consensual knowledge about relative clauses, is the distinction between the binding relation (between a relative determiner and the antecedent; I formalize this relation as Relative Binding) and the coreferential relation (between the relative phrase and its antecedent; I exclude this relation from the scope of sentence grammar).

Other aspects of the framework within which I investigate Bangla open clauses come from the analysis by Bresnan (1970) of the Complementizer node, which her use of bar notation identifies as the Specifier of S. In particular, one of my working hypotheses is an amplified version of Bresnan's claim that relative clauses and constituent questions indirectly, and complement clauses and yes-no questions directly, owe their gross grammatical properties to the behaviour of the Complementizer in relation to the +Open phrases, if any, in the clause. Bresnan's analysis rested on a pioneering paper by C.L. Baker (1970).

So much for historical antecedents. Let me state the findings now.

If certain definitions of Y (yes-no questions), R (relative clauses), C (complement clauses), and Q (constituent questions) are accepted, in particular for Bangla and in general for languages which are 'similar enough' to Bangla, then a generative analysis, using standard concepts and rule types, can be written which accounts for the (dis-)similarities of internal and external syntax and semantics among these clause types.

The definitions are as follows. Y requires yes or no answers; this definition has the fortuitous effect, in Bangla, of excluding indirect alternative questions. R does not, for the purposes of this study, include relative clauses which follow their antecedents, but only those which parade their antecedents. C does not, in this study, include instances of S which occupy the S_2 slot in $[_S\ S_1\ [_{CONJ}$ je$]\ S_2]$ structures, but only instances of S′ that contain a true Complementizer *je* capable of undergoing Complementizer Preposing and of co-occurring with Conjunctions. Q does not, in this volume, include echo questions, which can be recognized by their intonation.

A language is 'similar enough' to Bangla in the sense required if it has one or more systems of Relative Words and Interrogative Words. Thus, German, which has two systems of Relative Words (the d- set, which also does Demonstrative/Definite work, and the *w*- set, which also does Interrogative work), is similar enough, while Japanese, which has no set which functions as a set of Relative Words – no words which can be at all considered 'relative words' – is not similar enough.

The generative analysis which accounts for the (dis-)similarities among Y, R, C, and Q in Bangla postulates basically the following state of affairs. Although Bangla is a language which, to follow the prevalent movement metaphor, 'keeps' its open phrases (its relative phrases and its interrogative phrases) 'in place' instead of 'moving' them anywhere,

grammatical rules nevertheless establish a relation of construal in relative clauses and constituent questions between the COMP position and the open phrasal DET position(s) in the clause. The grammar also relates each open Specifier (COMP and/or DET) position within an Open Clause (the term 'Open Clause' covers Y, C, Q, and R) to one or more antecedents outside that clause which bind the open SPEC position; in the case of +INTERROGATIVE open specifier positions, it is a semantic zero that does this binding; for –INTERROGATIVE cases the binder must be a syntactically present antecedent.

This study has shown that the grammar briefly characterized above accounts for the data of Bangla better than an alternative grammar of complement clauses and relative clauses (offered by Jayanti Chattopadhyay) which derives R by a relativization transformation and C by a COMP insertion transformation. On the basis of this result, students of Bangla may now hope to build an overall generative analysis (or other post-structuralist[23] analysis) of Bangla which, perhaps after a decade or two of further research, will begin to do for the synchronic study of Bangla what S.K. Chatterji's *Origin and Development of the Bengali Language* does for its historical study. The currently available synchronic grammars, e.g. Chatterji (1968) and Ray et al. (1966), in the traditional framework and the structuralist framework, respectively, are far too sketchy to enable users to situate the structure of Modern Bangla among language structures of the world. They also do not build the concepts which will be needed for compiling a scientific dictionary of the language. The present effort, since it skeletally covers most of the grammatical structure of the language, at least proposes a system which can be tried out and corrected.

Thus, one reason why it is important to have carried out this study is that we can now hope to put together a relatively coherent account of Modern Bangla structure by remedying the inadequacies of the present analysis.

Still on the descriptive level, but over a wider range, the present study is important because it establishes for one language the claim that the –INTERROGATIVE +OPEN Complementizer word of a complement clause and the –INTERROGATIVE +OPEN Determiner word(s) of a relative clause share the grammatical feature bundle [–INT +OPEN]. This claim, to the extent that it is correct, offers a general account of the similarity between the C complementizer and the R determiner/pronoun system in a variety of languages (Sanskrit, German, French, Russian, etc.).

On a theoretical level, this study is important for at least two reasons. First, it demonstrates that the +OPEN element (often known as WH because of its manifestation in English) is logically a variable and not, as Chomsky and many others believe, a quantifier; this demonstration leads to the further suggestion that, as a semantic construct, an existential or universal quantifier is less basic than a variable, from which it follows that the interpretation of quantification must rest on the semantics of anaphoric binding relations. If this suggestion proves to be correct, linguistic semantics will cease to simply draw on the ideas of classical logic and will instead offer an analysis of what classical logic assumes to be unanalyzable primitives. Second, throughout this study, it becomes clear that the Strict Lexicalist approach to many empirical linguistic questions is more useful than the Mixed Lexicalist approach, Bresnan's approach more useful than Chomsky's. Thus, this study bears on one of the unresolved debates in linguistic theory.

[23] A 2020 addendum: when choosing this word in 1980, I was unaware of the Spivak-type use of the term post-structuralism.

Chapter 10

THE CHOMSKY-BRESNAN DEBATE

10.0 Differences of approach

Bresnan's theory of language, which I call Strict Lexicalism or SL, incorporates the thesis 'that semantic interpretation for natural language [need not] involve a translation into logical form' (Bresnan 1976a: 3n). She contrasts her research programme most sharply with 'the generative semantics programme [which assumed a need to relate] the syntactic categories of grammar to the categories of a variant of the predicate calculus' (1976a: 4). To an appreciable extent, the current Chomskyan research programme (which I will call Mixed Lexicalism, ML) for syntax and semantics, especially since *Questions of Form and Interpretation* and *Conditions on Rules of Grammar* (both now in Chomsky 1977a), has come closer and closer to 'generative semantics' so characterized. In particular, ML writings on the semantics of questions accept the view of many predicate logicians that one must posit a special interrogative quantifier — '?' in Hans Reichenbach's (1947) notation. For mixed lexicalists' acceptance of this view, see in particular Chomsky (1978), Chomsky & Lasnik (1977), and May (1977). The logical literature on questions is extensive. Belnap & Steele (1976) provide a 46-page bibliography. Karttunen (1977) and Boër (1978) provide Montague-grammatical alternatives to the Fregean logical accounts of interrogative semantics. Karttunen treats question words as quantificational. Boër's view corresponds to a particular SL suggestion by Brame, to whom we now turn.

10.1 The importance of Brame's suggestions

It is a truism that importance is in the eye of the problem-chooser. Generative grammarians today proceed on the assumption that for most purposes Chomsky's and Bresnan's contributions to their respective sides of the debate and the issues which they pinpoint as crucial are crucial, and that other people's writings and views are to be placed in relation to the cruxes as defined by the 'crucial' writings. It would be fatuous to dismiss such feelings as mythical and fashionable (even if they are to some extent both of these things), for they are the stuff that research energy is, perhaps unfortunately, often made of. There seems to be a need, in practice, for charismatic leaders in scientific inquiry.

Having said all this, though, I must point out that, for my purposes, the crucial thing about the Bresnan side of the debate is that someone on the SL side, Michael Brame, has come up with what I take to be particularly insightful proposals for the analysis of 'WH' words and for interrelating complement clauses and relative clauses. Brame has suggested (1979:120) that 'WH'-Words be regarded not as quantifiers but as positions for a bound element (if I understand him correctly), and (1979: 119) that complement clauses be regarded as, in effect, 'relative' clauses except that the whole clause (rather than some position in it) is 'relativized'. These proposals of Brame's, though they don't address mainstream issues, fit naturally into the SL theory of grammar which, stressing the role and complexity of the lexicon and of functional structure within propositions, looks for regularities involving tangible pieces of syntactic structure accessible to science.

In principle, someone on the ML side could have come up with similar ideas. It

certainly seems easy to translate the ideas into the ML framework as it stands at present; some linguists feel that to a large extent Chomsky's and Bresnan's conceptions are empirically indistinguishable 'notational variants' of the same theory. However, even if one grants this point as being at least partly valid, one must still ask which 'notational variant' is more worthwhile as a tool for specific sorts of investigation. A good discussion of an exactly analogous point is found in Fillmore (1977), whom I now quote. He is discussing criticisms

> that take the form, "I can do anything you can do" – the argument of MERE NOTATIONAL VARIANCE. ... This argument ... has something to say for it, of course. It is a version of the view that a change in a scientific paradigm should be tolerated only when the possibilities of the existing paradigm have been exhausted AND when a more satisfactory paradigm can be shown to exist. I sympathize with this view in general; my position, however, is that alternative paradigms – or even alternative notations – should be valued for the kinds of questions they force an analyst to ask. … Whether the answers to these questions can be WRITTEN DOWN within the terms of some other paradigm does not always seem important to somebody primarily concerned with whether the questions are important ones and whether they can be answered at all.
>
> (Fillmore 1977: 67-68)

Ever since Bresnan (1970) postulated a 'WH' element shared by interrogative and relative items and sought to explain cross-linguistic patterns involving these items, her work has tended in a direction which makes it easier and more natural to ask descriptive and theoretical questions about 'WH' items in her framework than in the (at first slightly, then increasingly) different framework of Chomsky and his associates, who have been more concerned, in theory and practice, with what they take to be movement rules and their effects in languages like English and French. So, as Fillmore says, even if it is possible to reformulate one's answers in ML language, the questions one can more easily ask in the SL language of Bresnan and Brame should continue to be pursued in that language.

The fact that it was Brame and not, say, Lasnik, who came up with the proposal that 'WH' words be regarded as bound rather than as quantifying and that complement clauses are clausally relativized clauses goes to show that Bresnan's way of doing linguistics makes it easier to ask worthwhile questions and propose fruitful answers, regardless of how much her differences from ML are a matter of theoretical style rather than empirical content.

However, there are some known empirical differences between the two approaches, especially in English grammar, but also in general linguistics. Let us work on some of these differences.

10.2 An interrogative quantifier?

I have been suggesting that it is possible to state Brame's proposals within ML. However, at least on the issue of the semantics of 'WH' words, the actual proposal of ML (May 1977, Chomsky & Lasnik 1977, Chomsky 1978, Fiengo 1978), contrary to that of Brame, is that there is an interrogative quantifier associated with the 'WH' element of English question words and that this quantifier binds a variable associated with the trace left by 'WH'-preposing. While

there is no reason to quarrel with the idea that the semantic element (whatever its character) associated with a 'WH' element binds a variable associated with 'the trace[24] of the WH phrase' (given that variable binding in linguistics, though not in classical logic, countenances situations where one variable binds another, since binding is a matter of anaphora), I do wish to criticize the Quantifier Hypothesis (QH) — the proposal that the semantic element associated with the 'WH' element is a quantifier of some sort.

First, Chomsky's acceptance of QH rests on (among other things) a misconception of the nature of QR. Second, his rule for interpreting 'WH' doesn't work. Third, even if one were to overlook these points, one would have to reject QH on empirical grounds.

Chomsky (1978: 6) states that 'movement of *wh*-phrases to COMP may be related to their quantifier-like character, hence a property of LF analogous to the structure-preserving property of NP-movement.' The context shows that he has May's (1977) Quantifier Rule QR in mind, a rule which, in course of the derivation of Logical Form (LF) from surface structure, preposes quantificational phrases like *every problem* and *some solution* and adjoins them to the left of S. What Chomsky is saying in the quote is that, just as NP-movement is a structure-preserving rule in that, when NPs move, they land in already base-generated empty NP sites, so also 'WH'-movement is a structure-preserving rule in the sense that a 'WH' phrase, when it is preposed, lands in a site to the left of S, exactly the place where, in LF, quantificational phrases belong. It is as though the 'WH'-preposing transformation were a syntactic anticipation of May's QR which applies in the post-syntactic component leading to LF.

However, Chomsky has based his suggestion on the incorrect assumption that May's QR is actually a preposing rule which adjoins material to the left of S. As May (1977: 47) is at pains to point out, 'since we have employed the notion c-command, which is defined over hierarchical structure, the linear sequence of the [quantified] noun phrase and the sentence from which it was extracted in logical form is irrelevant. Thus, we could just as well have

[24]Readers more familiar with recent writings on trace (due mostly to ML theorists) than with the earlier literature may see this as an acceptance of the ML position that there is a rule of 'WH'-preposing which, like other movement rules and unlike, say, rules of deletion, leave trace. But even in the recent literature, an author like Fiengo (1978) who basically accepts Chomsky's theoretical framework (rather than Bresnan's) regards as reasonable the alternative of base-generating empty MPs and coindexing them with other NPs, and for historical reasons continues to use the term 'trace' for such coindexed NPs. Thus, we see that the movement connotation of the term 'trace' is not absolute. The usage under which a node emptied by deletion under identity is called a trace has as much traditional warrant as the movement usage. Selkirk (1972), for empirical reasons, did not distinguish movement from deletion as a source of trace. Those empirical reasons, incidentally, still argue against the ML proposal (Chomsky & Lasnik 1977, Chomsky 1978) that deletion leaves no trace. Chomsky (1978:7) states that deletion under identity governed by subjacency derives *John went more often to Paris than Bill Ø to London*, where *Ø* is nodeless. In that case *John is going more often to Paris than Bill is Ø to London* also has no node between *Bill is* and *to*; so, *Bill is* here should be able to contract to *Bill's*. But *Bill's* is in fact unacceptable here. Thus, either it is movement and not deletion that Chomsky must postulate in this case (against the core assumptions of his current framework), or it is wrong to assume that deletion never leaves trace. Perhaps it is only deletion under identity that leaves trace.

Those who favour an ML approach might counter that another way to do it would be to have a base-generated null node sequence for *Ø* above. Fiengo (1978), for example, systematically uses 'zero allomorphs' instead of deletion processes. Provided that such node(s) survive in phonology, this alternative would in fact work – and be much closer to SL. But Chomsky (1978: 7) precludes this in his system by proposing a phonetic convention 'that $[_{\alpha}\ e]$ is automatically deleted unless α is indexed.' Thus Fiengo's alternative is unavailable to Chomsky, in this case.

represented the logical form [in terms of right-adjunction instead of left-adjunction]. In what follows we will represent quantified phrases in logical forms as standing to the left of the sentences from which they have been extracted, in order to maintain the graphic similarity of logical forms (in our sense) and standard logical representation.' If May is resorting to preposing rather than postposing merely as a matter of customary notation, surely the status of QR – a 'movement' rule for which no direction has been defined – differs materially from that of 'WH'-movement which, if it exists, is an inherently leftward movement rule. Therefore, Chomsky's conclusion about the 'property of LF analogous to the structure-preserving property' is ill-founded. To the extent that this consideration served to argue for or to reinforce the Quantifier Hypothesis, it is now seen to have no force.

Turning to Chomsky's rule of 'WH'-interpretation, quoted below, notice that this rule (Chomsky 1978: 6) works for *who* but not for *what* in *Who saw what?*

> (9) $[_{\bar{S}}[_{COMP}[_{\alpha i}$ wh-phrase$][_{COMP}$ $\pm$WH$]][_{S}\ldots[_{\alpha i}\ e]\ldots]]$
> […] In a structure such as (9) a rule of *wh*-interpretation will apply assigning the status of a quantifier to the *wh*-phrase in COMP and the status of a variable bound by this quantifier to an appropriately selected element within the position of $[_{\alpha i}\ e]$, the trace of *wh*-movement.

So, even in English, the language for which the rule was formulated, the rule obviously does not account for its data. Presumably, it will be necessary, in a careful ML account, to defer the interpretation of WH-phrases until QR has applied to adjoin all WH-phrases, regarded as quantified phrases, to S. But then (9) itself is no longer a fit structure for interpretation. The WH-phrase in (9) must first be moved into a position adjoined to S before it can be interpreted. In (9) the phrase is adjoined to COMP, not to S, if we use the term 'adjoin', as Chomsky does, in the sense of what Baker (1978) calls aunt-adjunction.

Quite apart from these objections to Chomsky's version of QH, the QH idea as such is suspect on empirical grounds.

My first argument against QH is based on Barbara Partee's (1975: 265) observation that (1-a) and (1-b) elicit considerably different judgments from different speakers, 'plus plenty of undecideds', as to whether the two sentences have the same range of readings in regard to quantifier scope.

(1) a. It seems that some man loves every woman
    b. Some man seems to love every woman

In contrast, (2-a) and (2-b) are uncontroversially distinct from each other with respect to what a proponent of QH would call the 'scope of the two interrogative quantifiers'.

(2) a. *ke kake* SaMtar kaTte dekhlo?
    '*Who* saw *whom* swimming?'
    b. *ke* dekhlo *ke* SaMtar kaTche?
    '*Who* saw *who* was swimming?'

Any theory incorporating QH, regardless of its stand on quantifier scope (regardless, therefore, of whether the theory predicts that *a* and *b* will have distinct readings, that they will not, or that

the answer is indeterminate), is expected to provide parallel answers to the questions that arise at (1) and at (2). Thus, any such theory is observationally inadequate, for (1) is, and (2) is not, controversial. Notice that the argument as I have worded it is independent of particular views on the role of VP or S.

A second argument against QH rests on lexical facts. The meanings of universal and existential quantification are idiosyncratically built into lexical items of various syntactic categories. For example, *per*, a particle, *apiece*, an adverb (a particle?), *entire* and *whole*, adjectives, contain universal quantification in their lexical meaning, without displaying any morpheme or any morpheme combination uniquely associated with universal quantification. See McCawley (1977) for an account of the phenomenon. But the languages I am acquainted with feature an identifiable question; and, more importantly, the 'question quantifier' shows no cases corresponding to *per* for the universal quantifier – a unimorphic lexical item that inextricably fuses the alleged quantification with other material into its semantics.

The contrast is not with universal quantification alone; languages known to me do not show any characteristic 'existential' morpheme or morpheme-combination either. It would be a mistake to see my work on Bangla indefinite forms as showing that *o* is such a morpheme; the point is that many things other than *o* express existential quantification, whereas *k* uniquely indicates interrogatives.

A third argument against QH, again lexical, has to do with subcategorization. Predicates and other governing elements subcategorize for questions. To illustrate from English, the predicate *wonder* and the composite preposition *as to* govern interrogative clauses and no declarative clauses. However, no predicate or governing element is known to subcategorize for universal or existential quantification or for clauses containing such quantification. This would indicate that the latter is either irrelevant to sentence grammar or far more tangentially relevant than interrogation. Therefore, it is theoretically dubious to classify interrogatives together with existentials and universals as 'quantificational'.

It might be thought that it is incumbent on me to state and to disprove at least a large body of existing arguments to the effect that questions are best understood in terms of quantification. This task is rendered difficult by the fact that May (1977), the most thorough ML treatment yet of quantifiers, simply presents a theory which includes QH as a claim, and argues for the theory without polemic against any non-QH theory. It is difficult to pinpoint specific arguments for QH. One such argument was recently presented by Chomsky in his November 1978 lectures at Columbia University. The original discussion of the data on which the argument rests appears in *Conditions on Rules of Grammar* in Chomsky (1977a); see especially (105), the principle that 'a variable cannot be the antecedent of a pronoun to its left', which Chomsky in his lectures called the Leftness Condition. In the light of this principle Chomsky examines 'logical forms' such as the following.

(3) a. [$_S$ someone [$_S$ the woman he loved betrayed t]]
b. [$_S$ someone [$_S$ t was betrayed t by the woman he loved]]

(4) a. [$_S$ Who [$_S$ did the woman he loved betray t]]
b. [$_S$ Who [$_S$ t was betrayed t by the woman he loved]]

I have shown the forms as May's QR would derive them and have written *t* only for traces of the phrase left-adjoined to S. We are concerned with possible anaphora between *he* and *t*. The

Leftness Condition, as Chomsky states it, correctly predicts (for some dialects) that a reading where *he* is anaphoric to *t* should be unavailable for *a* but not for *b*. Now consider (5).

(5) a. The woman he loved betrayed John
b. John was betrayed by the woman he loved

Provided that stress falls on *betrayed* rather than (contrastively) on *John*, both the sentences in (5) allow *he* to be anaphoric to *John*, indicating that it would be inappropriate to design a 'logical form' for (5) where *John* was QR-moved to a position in adjunction to S. These facts enable Chomsky to conclude that, contrary to claims made by Montague grammarians working on question (Chomsky does not refer to Karttunen or Boër), question words are not like proper names. He adds that the facts also justify regarding question words as quantifiers.

Notice, though, that QH plays no role in the argument. The crucial fact about (3) and (4) is that the *t* nearest to *he* is to its left in the *b* sentence in each case and to the right in the *a* sentence. This *t* is interpreted as a variable, and the Leftness Condition applies to yield the correct predictions. Whether the *t* is specifically the trace of a **quantifier** expression is irrelevant. In the *b* sentences, in fact, the relevant trace is the one after *betrayed*, and it is the trace, not of a quantifier, but of the subject *t* before *was*. Thus, instead of structure (4), which reflects QH, one might as well use (6) – the surface structure in Chomsky's system.

(6) a. $[_{S'}\,[_{COMP}\,[_{NP}$ who$][_{COMP}$ +WH$]][_{S}$ did the woman he loved betray t$]]$
b. $[_{S'}\,[_{COMP}\,[_{NP}$ who$][_{COMP}$ +WH$]][_{S}$ t was betrayed t by the woman he loved$]]$

This structure, which does not depend on QH, enables Chomsky to make the same predictions. Thus nothing follows, from the considerations presented by Chomsky, about whether QH is a quantifier element.

To pursue the matter further, bearing in mind Chomsky & Lasnik's (1977) claim that relative 'WH'-Words are not quantifiers, consider (7).

(7) a. the man who the woman he loved betrayed t
b. the man who t was betrayed t by the woman he loved

The pattern mimics that of (3) and (6). By Chomsky's reasoning it should follow, contrary to Chomsky & Lasnik's express claim, that relative *who* is a quantified phrase as well. Therefore, either, as I maintain, the reasoning is faulty, or one must conclude that the 'WH' element is a quantifier in relative clauses as well (the latter is the stand taken by someone who shares many of Chomsky's views in this debate, Robert Fiengo (personal communication)). Consider, for argument's sake, the option of adopting Fiengo's view. Suppose 'WH' is indeed always a quantifier, contrary to Chomsky's (1978: 28) proposal that in a sentence like *I asked the man who you had visited* (on the relative reading) the word *who* 'is devoid of semantic content'.[25]

[25] This proposal seems at odds with the proposal (Chomsky 1978: 31) that a cleft sentence has the structure *it is* α S, where α = NP or PP, and that there is a 'rule of interpretation associating α with the COMP of the embedded sentence'. How does a rule of interpretation associate a phrase with something devoid of semantic content? However, the contradiction may be only apparent. Perhaps this particular 'rule of interpretation' is a rule of 'construal' in the sense of chapter 7. In that case COMP not only can but must be considered 'devoid of semantic content'.

As Fiengo has pointed out to me, the claim that 'WH' is a quantifier and not a variable entails that nothing can bind 'WH'. But consider the promotion or raising analysis of relative structures, which is as widely accepted as any other for French and to a lesser degree for English. This analysis 'promotes' or 'raises' an antecedent phrase into an NP position to the left of S′, leaving a 'WH'-marked trace which then preposes into the COMP of that S′. The classical ML conception of binding as a necessary consequence of movement unambiguously requires that, in such an analysis, the relative 'WH'-phrase be considered 'bound' by the antecedent. Chomsky (1978), one of the primary documents of ML, assumes that raising is a reasonable, if not the reasonable, analysis of English relatives (cf. p. 31). Thus, ML at present does not allow a coherent formulation of the hypothesis that 'WH' in relatives is a quantifier. If ML is modified (in particular, by dropping or radically altering the raising analysis of relatives), it may become possible to view the relative-antecedent relation in terms of some notion other than binding, and thus to salvage the idea that 'WH' might be a quantifier in all of its uses. I can envisage this prospect, but cannot discuss it, in the absence of specific proposals. For the time being, it seems that key assumptions of ML make it impossible to regard relative 'WH' as a quantifier, and that therefore the parallelism between (7) and (6) shows that interrogative 'WH' cannot be a quantifier either, or at least that Chomsky's argument to this effect does not work.

It is difficult to give specific arguments showing that Brame's alternative – of regarding 'WH' as always a bound element – must be adopted. But it can be shown that the alternative works for the data at hand. Returning to the surface structures (6) reproduced below in string form, notice that in the *b* sentence the variable which plays antecedent to *he* is associated with the second trace, which is itself bound by an antecedent, the variable associated with the first *t* in *b*.

(6) a. who +WH did the woman he loved betray t
b. who +WH was betrayed t by the woman he loved

Thus, it is unproblematic to propose further that the first *t* in *b* in turn is a bound variable whose antecedent is associated with *who* and is itself a bound variable, and that the antecedent which binds the latter is a semantic zero element in (6) and the overt phrase *the man* in (7). Moreover, notice that, in (7), the c-command condition is satisfied. These considerations by themselves show only the possibility, not the necessity, of postulating variable status for 'WH'. Positive arguments become available if one accepts a raising analysis for relatives, or the overall structure of my Bangla grammar (which makes crucial use of the idea of Open Words being variables – recall that Open Words is my term for what most of the English-centred literature calls 'WH' Words), including in particular my analysis of indefinite expressions.

(7) a. the man who –WH the woman he loved betrayed t
b. the man who –WH t was betrayed t by the woman he loved

10.3 English infinitives – clausal and phrasal

One of the empirical issues in the ML-SL debate is the status of the English infinitival system. Bresnan (1973) reiterated the hypothesis of Lees (1960) and Emonds (1970) that many

infinitival complements in English were in fact VP and not S. At no stage did she make the comprehensive claim that all infinitivals in English were VP. At least for Bresnan, the existence of *It is impossible for there to be thirty people in that hole* demonstrates that some infinitivals must be clausal. In contrast, Chomsky (1978) does make the comprehensive claim that all English infinitivals are S′. Associated with this is Chomsky's (1977b) reanalysis of *easy-to-please* structures in terms of 'WH'-movement. Thus, ML now claims that *John is easy to please* and *John is an easy person to please* derive from *John is easy who(m) to please* and *John is an easy person who(m) to please*, respectively, by obligatory deletion of *who(m)*. The SL position, in contrast, is that the underlying structures are *John is easy to please (John)* and *John is an easy person to please (John)*, respectively, where the parenthesized occurrences of *John* are present in some and absent in other versions of SL.

In a sense, this issue does not lie at the heart of the debate, since the claim of each side translates into the framework of the other. (This is an unguarded and possibly inaccurate statement. To translate the ML claim into SL one would need to introduce extra devices of a sort that no SL theorist, perhaps, would accept today.) However, the reason why strict lexicalists favour the analysis they do stems from the foundations of SL reasoning. If some infinitival complements are phrasal, a gap gets filled in the subcategorization system, for then VP stops being a category with the unique property that matrix words cannot subcategorize for it.[26] The link between the philosophy of ML and its adoption of the clausal analysis is harder to localize, but, roughly, their claim manifests a typical proneness for the postulation of empty syntactic nodes to lighten the burden on the system of logical form.

Brame (1976, 1977) gives several cogent arguments supporting the SL against the ML position. To these I would like to add a few further considerations.

One argument against the claim that all easy-to-please infinitivals are 'WH'-clauses is based on the following English sentences.

(8) a. John is an impossible person to discuss it with Ø
b. *John is an impossible person with whom to discuss it
c. *John is an impossible person whom to discuss it with

According to the ML analysis, (8-a) derives from (8-c) via deletion of *whom*. But if (8-c) is a well-formed underlying structure, (8-b) should derive from (8-c), and yet (8-b) is ill-formed as a surface structure for many speakers, although no independent reason for its ill-formedness has been adduced. I conclude that (8-c) does not exist as an underlying structure and that (8-a) must therefore derive in some other fashion, perhaps as in SL.

My first argument against the claim that all infinitivals are clauses is based on the following data.

(9) a. They saw *the boat capsize*
b. They let *the honey drip*
c. They made *the baby scream*
d. They felt *the ions collide*

[26] Mark Baltin objects to this point on the basis of analyses in which *have* and *be*, as auxiliary verbs, subcategorize for VP. But the point is about the regular 'lexicon' which, presumably, does not include function words like *have* and *be*. So, these cannot be cited as counterexamples.

e. They watched *the sulker sulk*
f. They heard *the bather sing*
g. They helped *the thief escape*
h. They had *us do the dishes*

(10) a. *(*For) the boat capsize would be a disaster: *(*X) = * with or without X
b. *(*For) the honey drip would be a disaster
c. *(*For) the baby scream would be a disaster
d. *(*For) the ions collide would be a disaster
e. *(*For) the sulker sulk would be a disaster
f. *(*For) the bather sing would be a disaster
g. *(*For) the thief escape would be a disaster (OK with *for* if escape is an N)
h. *(*For) us do the dishes would be a disaster

The ML position regarding the italicized parts of (9) is that they are embedded sentences (Chomsky 1977a). Therefore, ML syntax predicts that (10) should exist, which is not true. Therefore, the initial idea about (9) is wrong. The italicized sequences are not sentences, but NP VP sequences.

My second argument against the view that all infinitivals are clausal is based on data which Dougherty (1969) uses in the context of a different argument. He assumed the conclusion I am arguing for here – that not all infinitivals are clausal in surface structure. Recall that this conclusion is found as early as Lees (1960), the first study of English within the generative tradition, and that the idea of all infinitivals being clausal is quite new and undemonstrated.

(11) a. John and Bill wanted Mary and Sue to go to New York and Chicago respectively
b. John and Bill wanted to go to New York and Chicago respectively

Dougherty shows that a satisfactory account of the range of readings for these sentences hinges on the assumption that at surface structure (11-b) is a simplex clause, while (11-a) contains an embedded clause. If one accepts his analysis, then, given the absence of any account of the data consistent with the ML claim that all infinitivals (including the one in (11-b)) are clausal, the facts about (11-a) and (11-b) underpin an argument supporting the SL against the ML analysis.

This argument needs fortification, for a mixed lexicalist might choose to retort that the interpretation of *respectively*-sentences involves some sort of Minimal Distance Principle (MDP), perhaps along the lines of Chomsky (1978). This retort can be quickly dismissed if one assumes that the subject of *go* in (11-b) (a mixed lexicalist would assume that the infinitival must be an S′ in surface structure and therefore has a subject) is PRO at surface structure, for then one clearly sees that MDP in conjunction with that surface structure makes the wrong predictions. But if one assumes instead that the surface structure of (11-b) arises by Reflexive Deletion, as in Chomsky & Lasnik (1977), then indeed ML can, with some effort (i.e. by allowing the relevant interpretive rule to see the output of deletion), replicate Dougherty's results; it then becomes necessary for me to argue against MDP, since on such assumptions (11-a, b) become neutral examples, incapable of helping us to choose between ML and SL.

10.4 The Minimal Distance Principle

Chomsky (1978) proposes that, other things being equal, an anaphor must be assigned the nearest eligible antecedent. One example where other things are not equal is the surface structure (in Chomsky's system) *John promised Bill PRO to leave*, where the verb *promise* assigns subject control and therefore the antecedent of PRO is *John* rather than – as MDP would have predicted – *Bill*.

In the light of this proposal, consider the following data.

(12) a. I acquired from my teacher an ability PRO to speak Swahili
b. My teacher imparted to me an ability PRO to speak Swahili
c. I received from my teacher a promise PRO to speak Swahili
d. My teacher made me a promise PRO to speak Swahili

(13) Bill and Mary exchanged promises PRO to write

These sentences[27], represented in the form of ML surface structures with PRO, contain no verbs of obligatory control in the sense of Chomsky & Lasnik (1977). Thus 'other things' should count as being 'equal', and MDP should apply to yield anaphora between PRO and *my teacher* in (12-a) and (12-c), between PRO and *me* in (12-b) and (12-d), and between PRO and *Bill and Mary* in (13). The prediction fails. The antecedent is *me* in (12-a) and (12-b), *my teacher* in (12-c) and (12-d), and individually *Bill* and *Mary* in (13).

One cannot solve the problem by lexically specifying *ability* and *promise* as assigning 'subject control', 'complement control', and the like, since these nouns, in these sentences, have no syntactic subjects.

It seems that the sets of thematic relations associated with these sentences determine the appropriate antecedent for PRO in each case, and, in (13), determine that PRO has no unitary antecedent, but that Bill promised Mary that he would write and Mary promised Bill that she would write. It is clear that MDP is not the correct principle. In view of this refutation of MDP, my argument based on Dougherty's data supporting the SL against the ML position on infinitives is strengthened, since ML no longer has the option of resorting to MDP as an alternative explanation for Dougherty's facts.

10.5 Passives

One of the major issues in the SL-ML debate is the existence of rules moving NP, in particular, the passive preposing rule (or, in some versions of ML, applications of some general movement rule which have the effect of passive preposing). Strict Lexicalists hold that passive structures are base-generated; Mixed Lexicalists maintain that they are derived by transformation. Again, I feel that Strict Lexicalists have already amply justified their position, and wish simply to add new arguments in support of SL.

[27] I am setting aside grammarians' nightmares like *To list their hopes in order of seniority, the three princes hoped PRO to live long, PRO to love widely, and PRO to serve diligently.*

My first new argument against deriving passives transformationally is based on the fact that a grammar of English which base-generates all passives – those with the auxiliary *be* as well as those with *get* – can approach the lexical relationship between transitive *get* and *have* on the one hand and intransitive *get* and *be* on the other as a unitary relationship. This relationship, which I will call TIR (the Transitive-Intransitive Relation), occurs between, for example, *The project got going* and *They got the project going*, and between *They had three projects under way* and *Three projects were under way*. Any description of English, whatever its approach to the passive, will need to accommodate TIR as a lexical generalization with syntactic consequences because of cases like these. In this light, consider now *The peaceniks got bashed up by the cops* and *The finks got the peaceniks bashed up by the cops*, or *The peaceniks were bashed up* and *The finks had the peaceniks bashed up*. Clearly the same generalization obtains, and to express it one must forgo the pleasure of deriving passives by transformation.

SL also has a natural explanation for the following facts which ML cannot explain in a principled fashion.

(9) a. They saw the boat capsize
b. They let the honey drip
c. They made the baby scream
d. They felt the ions collide
e. They watched the sulker sulk
f. They heard the bather sing
g. They helped the thief escape
h. They had us do the dishes

(14) a. *The boat was seen capsize
b. *The honey was let drip
c. *The baby was made scream
d. *The ions were felt collide
e. *The sulker was watched sulk
f. * The bather was heard sing
g. *The thief was helped escape
h. *We were had do the dishes

The SL explanation is that a general redundancy rule marks all verb subentries with a subcategorization for infinitives of this sort (for infinitives without *to*) as not occurring in the passive frame. A mixed lexicalist could in principle co-opt this explanation by postulating surface lexicalization. I will show later in this section that such a move would run counter to Chomsky's favourite reason for deriving the passive transformationally, and in section 10.6 that the route is blocked anyway since there is Bangla evidence against surface lexicalization in general. Alternatively, a mixed lexicalist can try to find principles (such as conditions on rules or on structures) which block (14). If such principles do not differentiate infinitives with *to* from those without *to*, the outcome will be the incorrect exclusion of well-formed sentences like *The thief was helped to escape* and *The baby was made to scream*. If the conditions do pick out *to*-less infinitives as the relevant environment, then there is the further problem that, for many speakers, *Who did they watch sulk?, What did they see capsize?*, and the like are well-

formed, so that a prohibition against an empty subject of a *to*-less infinitive will not do.

Norbert Hornstein (personal communication) has suggested that a *to*-less infinitival S′ is a propositional island in the enlarged sense indicated in Chomsky (1977b). The trouble with invoking the Propositional Island Condition becomes especially clear when we substitute for it Chomsky's (1978) Nominative Island Condition which superseded it: no Nominative anaphors may be free in S′. In current ML proposals, PRO carries no Case; there is thus no mechanism available that would get PRO to bear Nominative Case (as required if this condition is to do the job) in **The thief was helped PRO escape* but not in the grammatical string *The thief was helped PRO to escape*. If we decide that infinitives without *to* really have an occurrence of Tense in front of the V and that this Tense assigns Nominative across the board, then we will undesirably generate **They had we do the dishes* and the like. Even apart from the specific problems associated with the Nominative Island Condition, the Propositional Island Condition (originally the Tensed Sentence Condition) will need to distinguish (15) from (16) and thus must function within a theory which regards *to*-less infinitives, contrary to Hornstein's suggestion, as non-propositional.

(15) a. Why are they letting the honey drip on each other's feet?
b. Why are they letting the baby fall on each other's laps?

(16) a. *Why do they permit that the honey drip on each other's feet?
b. *Why do they permit that the baby fall on each other's laps?

The only remaining possibility for ML is that Opacity, originally the Specified Subject Condition, might block (14).

The Specified Subject Condition prohibited rules from relating positions on either side of a Specified Subject (the exact meaning of Specified in that earlier system does not matter here). But the NP movement application in question here relates a Subject position to another position; so, that version of the Condition does not apply. As for Opacity, which states that if an anaphor A is c-commanded by the Subject of phrase B, B minimal, then A cannot be free in B, examination of any of the examples in (14) – e.g. (14h), *we were had t do the dishes* – will show that there is no anaphor, free or otherwise, c-commanded by the Subject of *t do the dishes*, so that even this version of the Condition has no bearing. Thus ML is left without a principled explanation for the ill-formedness of (14).

In order to co-opt the SL explanation of the phenomenon a mixed lexicalist will need the hypothesis ('surface lexicalization') that lexical insertion takes place in surface structure. In section 10.6 I argue against this hypothesis. For the moment I will show that resorting to surface lexicalization for this purpose would undermine Chomsky's principal argument for the ML account of passives. In his November 1978 Columbia University lectures Chomsky cited as the most persuasive (though indirect) piece of evidence for his position the fact that English idioms show an asymmetry which one would not expect if passives were base-generated. To wit, there are active-only idioms like *kick the bucket* (while *John kicked the bucket* can have the idiomatic reading 'John died', *The bucket was kicked by John* has only the literal reading), and there are active/passive idioms like *keep tabs on* (*Bill kept tabs on John* and *Tabs were kept on John* are equally idiomatic), but there are no passive-only idioms. While SL can only 'describe' the absence of passive-only idioms – by not having any in the lexicon – ML, Chomsky argued, has an explanation for it. The explanation is that idioms are lexical items which get inserted

into base-generated structures, and passives are not base-generated structures. Now, if this argument is accepted, surface lexicalization must be rejected as a way out of the difficulty connected with (14), since it is crucial that lexicalization occur at deep structure.

I am not, of course, accepting either Chomsky's argument or surface lexicalization, but merely pointing out that the ideas are incompatible. Lest my rejection of surface lexicalization in section 10.6, coupled with the foregoing discussion, leave the impression that I find Chomsky's argument persuasive, let me specifically controvert the latter. *He was hard put to it to hold back the army* (Tolkien, *Silmarillion*, p.193), *The point is well taken*, and (I am indebted to Joanne Grumet for the following example) *The die is cast* are sentences of the sort needed to falsify his theory. Since they are well-formed, and since they have no active counterparts, his theory is, in fact, inconsistent with the facts of English.

10.6 Lexical insertion in surface structures

Several mixed lexicalists currently favour the hypothesis that lexical insertion takes place at surface structure. Fiengo (1978) makes particularly consistent use of it.

I have found an argument against this hypothesis in the Bangla verb system.

Both a compound verb and an infinitive-modal sequence in Bangla have the surface constituent structure $[_V$ V V]. But the sets of vector and modal verbs overlap, posing a major problem for surface lexical insertion (a vector verb is the second V in a $[_V$ V V] construction, in what is known as a 'compound verb' construction). At deep structure, however, the constructions differ in constituent structure. The compound verb has the deep structure $[_V$ V V]. But the derivation of the modal construction, as shown earlier in this volume, has as [... $[\ldots V]_{VP}$ V]$_S$ as its starting point. A reanalysis rule applies to produce the compound-verb-like surface structure. Thus, lexical insertion at deep structure can easily distinguish the two constructions. Quite naturally, no lexical subcategorization relations hold between particular infinitive verbs and their modals, whereas such relations do exist (see Dasgupta (1977) for a study) between the pole and the vector of a compound verb. (The pole is the first V and the vector the second V.)

Arguments known to me which purport to show a need for surface lexicalization deal with facts which could also be handled by not allowing syntactic transformations to interfere with word morphology – not introducing Case markers or Case-governing elements transformationally, for example. I will not deal with the arguments case by case.

10.7 Thematic relations and anaphora

Chomsky in his 1978 Columbia lectures distinguished the 'system of syntactic computation' of natural languages from their 'conceptual system' and suggested that they stem from different faculties of the mind and interpenetrate in the lexicon. He added that thematic relational notions like Agent seemed to pertain to the conceptual and not to the computational system.

From this viewpoint one may usefully see the difference of approach between ML and SL in terms of preferred emphasis – on the computational side in the former and on the conceptual side in the latter. We note, for example, the growth of detailed mechanisms for interpreting the functional/relational structure of propositions in recent SL work, in contrast to

the growth of mechanisms of anaphoric interrelation often involving abstract empty nodes in recent ML work. It was no accident that, when attacking MDP in section 10.4, I swung towards a solution in terms of thematic relations. I will now recall data which Chomsky cursorily examined during the formative period of ML, data indicating that the very basic notion of Opacity (then the Specified Subject Condition) rested on arbitrary exclusion of relevant facts.

(17) Why are John and Mary letting the honey drip on each other's feet?

(18) *Why are John and Mary letting Bill drip honey on each other's feet?

(19) Why are John and Mary letting the baby fall on each other's laps?

(20) *Why are John and Mary letting Bill drop the baby on each other's laps?

Chomsky (1977a: 124) notes these facts and mentions the possibility, which he chooses not to pursue, that the notion of Agent rather than Subject might be the appropriate one. It has remained unclear why that possibility was felt to be less worth pursuing, why movement and anaphora should be the privileged objects of study in ML linguistics.

Research on anaphora which uses referential indices as in ML is based on the hope that there is a well-defined phenomenon of anaphora which can be profitably formalized as index-binding. That that hope may be less than justified becomes apparent vis-à-vis the use versus mention problem pointed out by John Grinder in 1971:

(21) It is clear why people call Acid Eddie that

(22) Acid Eddie says that he is always tripping

(23) *It is clear why people call Acid Eddie him

(24) *Acid Eddie says that that is always tripping

(25) Even Acid Eddie's mother calls him that

Grinder states, 'Presumably the referential indices (adopting Chomsky's convention) will be of use in separating the antecedent/anaphor possibilities.' While the use of indices has grown in complexity and quantity, no progress has been made in solving the problem posed by (21)-(25). Perhaps the use of *that* in the above cases can be explained in terms of a theory of mention along the lines of Partee (1973), but in the absence of such an explanation we must bear Grinder's facts in mind as poorly understood counterexamples to theories of anaphora which, like ML work, use simple-minded (though often operationally complex) devices to indicate shared and disjoint reference.

10.8 Verb phrases in diachrony

An SL approach to English, being sensitive to complexities of the sort typically registered in

lexical mechanisms, is also better equipped to reflect linguistic change. For example, the SL position that English infinitive structures are phrasal in some sentences and clausal in others is in tune with the fact, acknowledged by Chomsky & Lasnik (1977) but not expressed in their analysis, that English is changing from a stage where all infinitives were phrasal towards a stage where, perhaps, all infinitives will be clausal. If we assume, naively but perhaps not far from the truth, that German resembles the older stage of English in this respect, then it is significant that Manfred Kohrt (1975) has shown, in effect, that German infinitives with *zu* (cognate to English *to*) are phrasal rather than clausal. Bangla as we have analyzed it is like German in this respect.

Perhaps the most significant clausal property of the newly emerging present-day English infinitive structure is its ability to be questioned and relativized, as in *It is unclear who to see, a knife with which to cut diamonds*. As Chomsky (1978) points out, it would be cumbersome to extend the VP system to accommodate these expressions within a new PS expansion introducing a COMP node to the left of a VP or VP′. These structures are best seen as clauses whose verb is non-finite. The fact that Bangla lacks such structures seems to me to mean that Bangla infinitives are purely phrasal, never clausal. Note also that those English infinitives which are most clearly non-clausal (in that they do not occur in subject position), the infinitives without *to*, are also unable to support questioning or relativization: **It is unclear who see, *a knife with which cut diamonds*. The only exception to this remark is a formulaic construction with *why* (*Why try?*, *Why bother?*) which never takes *to* (**Why to try?*), and cannot embed (**I wonder why bother*).

First of all, the explicit SL position that some infinitives are VP and others S′ in English makes it possible to begin to understand the details of this linguistic change in English. Secondly, SL, with its claim that Open ('WH') elements are variables, and with its conception of the special relation between COMP and Open elements (first formulated as the Baker-Bresnan universal, but generalizable, as in the present inquiry, into the idea that some sort of construal must link the Open phrases in a clause to the complementizer node of that clause), enables us to hypothesize that to be a clause is to be able to take a complementizer and thus to enter the realm of interrogation and relativization, just as to be an NP or AP is to take a determiner, which confers similar privileges, since NP, AP, and S′ are the hypophoric nodes (the nodes which can have phoric specifiers). This hypothesis is the beginning of a deeper understanding of, among other things, the shift in English from a VP infinitive system to an S′ infinitive system.

Let me pursue this idea (that to be a clause is to be able to be open) by looking at Openness from a broader point of view.

10.9 Pre-terminal words

It seems to me useful to regard the contents of COMP and DET in Open Clauses and Open Phrases as inflectional markers. On this view, the Open/Non-Open distinction and Interrogative/Non-Interrogative distinction become inflectional dimensions of the same order as Tense and Case. My main reason for making this suggestion is that if it is accepted then the conjecture of section 2.2.1.1 can link up with the rest of the grammar. The conjecture, that every pronoun is 'grammatically nothing but a set of values for various pronoun-relevant dimensions', reinterpreted in the light of findings in later chapters which make it possible to

localize many of the ‘pronouns’ in DET (rather than N or NP), is intrinsically interesting because, if true, it gives content to the widespread feeling that pronouns are somehow a non-lexical category. But it is also of extrinsic interest because of some questions it forces investigators to ask.

One such question: are the Dimensions of Openness and Interrogativity to be regarded as conjugational, declensional, or neither?

Although the initial, conservative temptation is to say ‘neither’, it may turn out to be more fruitful to say ‘declensional’, an answer that obviously involves not confining ourselves to the narrow sense familiar from Latin and Sanskrit grammar. We must explore the links with grammatical Term relations (I am invoking the notion ‘Term Relation’ current in Relation Grammar: Subjects and Objects count as ‘Terms’) with some reference, where needed, to thematic relations.

Obviously, it is not literally true that the relative versus interrogative distinction for DET has something to do with the ‘Case’ properties of the relevant NP or AP. Since the construal system ensures that the processing of an Open DET always goes through an Open COMP, the focus of our question should be on the clause. It is the clause which ‘inflects’ along the dimensions of Openness, Interrogativity, and ‘Phrasality’ (to give a name to the Dimension on which one point is ‘DET-Open’ or ‘Phrasally Open’ and the other point is ‘COMP-Open’ or ‘Clausally Open’), although these inflectional markers are located in the Specifiers of various hypophoric nodes (including the S′ node itself).

I find it useful to regard an Open Clause as analogous to a Genitive NP along the following lines: an Open Clause, like a Genitive NP, subordinates itself to something else (one or more Antecedents in the Clause’s case, corresponding to a possessor for the Genitive NP), and in doing so approaches the behaviour of a slightly different category which is inherently subordinate (the Open Clause approaches the behaviour of a Participle Phrase; the Genitive NP approaches the behaviour of an Adjective Phrase – quite strikingly in the phenomena examined in section 10.10), thus departing slightly from the more typical behaviour of the original category which is inherently non-subordinate (the Open Clause is slightly less of a clause in that a ‘prototypical’ clause is a Declarative Main S′; the Genitive NP is slightly less of an NP in that a ‘prototypical’ NP bears a Term-relation to the verb in the clause or VP). If this analogy is accepted, it may be expressed by saying that Openness is a declensional Dimension along which a Clause declines, and that Open is for a Clause what Genitive is for a Noun Phrase.

There is empirical evidence for something like the ‘clauses are Case-declined’ contention in the Japanese, Dyirbal, and Gumbainggar data mentioned in chapter 9. Since those data patterns crucially involved Conjunctions, it is clear that, pending investigation of the Conjunction systems of many languages, several fundamental issues surrounding the Chomsky-Bresnan debate will remain unresolved, as will other questions in this domain.

As the last theoretical act of this inquiry, let me note that the decision to regard Openness as a declensional Dimension of S′ and the occupants of Open DET and COMP nodes as expressing values on this Dimension raises the question of lexical entries quite sharply. One does not normally see markers of declensional Dimension values as entries in a lexicon. How, then, do the contents of Open DET and COMP nodes come to occupy these nodes?

As part of our thinking about this question we should consider the status of these nodes. Bar-notational research regards them as Specifiers of AP/NP and S′, respectively. Surely the notion ‘Specifier’ is a functional, not a categorial, notion, in the sense of ‘functional’ and ‘categorial’ outlined in Chomsky's (1965) discussion of the difference between notions like

'Subject' and notions like 'Noun Phrase'. An NP can play various roles – Direct Object, Subject of an NP, etc. So, the category NP must be distinguished from the roles it plays. But what can a Specifier of S′ do but specify S′? What can a Conjunction do but conjoin? Categories such as Conjunction and Specifier-of-S′ (the term Complementizer is just an informal abbreviation for Specifier-of-S′) are not, then, categories in the full-blooded sense, as distinct from functions. It is perhaps not enough to recognize this difference by calling them 'functional categories' in contrast to 'lexical categories' like N or V. Maybe the very processes whereby 'functional categories' obtain their terminal string content need to differ from lexical insertion. The 'Inflection Rule' idea which I have used earlier in this volume may need to be formalized more carefully and enlarged in scope so that it covers DET and COMP 'insertion' as well – and, when we come to understand conjunctions better, CONJ 'insertion'.

Speaking of bar notation, there is a notion intermediate between Specifier and Head which appears in Hornstein's (1975) formulation of bar notation – the notion of Modifier, which Hornstein does not explain, apparently because he assumes that the traditional grammatical definition of the notion carries over to his use of it. A typical Modifier category is AP. This volume has so far made no contribution to the task of understanding this area of bar notation. As a token contribution, let me propose that, at least in this one case, the Modifier-type node AP is an 'inherently subordinate' counterpart to the Head-type node NP, where subordination is what happens if you go Open (if you are a Clause) or if you go Genitive (if you are an NP). The following section studies the relation between some APs and some genitive NPs, partly in elucidation of the above remark, partly to clarify the analogy on which I based the conjecture that Open is the Genitive of a Clause, and partly to end this investigation on a note of looking at facts that lead into future work by making a problem precise.

Some forms in Bangla which end in *-r* are not clearly Genitive Nouns; they may be Adjectives. I here reproduce, unemended, a research fragment about these forms which was written in December 1977.

10.10 Terminal words: genitive nouns and adjectives

(26) i. N → N (CASE)
ii. CASE → GENITIVE, ...

Bangla Genitive Case suffixes – namely, *(e)r* and *kar/ker* – behave complexly. Noun stems that take these suffixes fall into three classes, which I will call M, L, and K.

M contains the vast majority of noun stems, like *kopi* 'cabbage/ cauliflower', *raja* 'king', *jiniS* 'thing', *ghOr* 'room', etc. They take *er* if they end in a non-vowel or in *o* from /Ø/ or have one syllable; they take *r* otherwise; henceforth I shall refer to taking *r* or *er* as 'taking *er*'.

K has less than ten members: *aj* 'today', *kal* 'yesterday/tomorrow', *kOkhon* 'when?', and some others. A member of K takes *ker* if it is either *aj* or *kal*; it takes *kar* otherwise (henceforth I shall refer to this behaviour as 'taking *kar*').

L, the interesting class, includes about a dozen stems such as *bhitor* 'interior', *opor* 'top', etc. Every L stem allows free variation between the suffixes *kar* and *er*. In some environments, however, they only occur with *er*. This set of environments, which I will call set A, tolerates no *kar* forms (to highlight the contrast in my glosses, I write 'GEN2' for *kar*):

(27) ghOrer bhitorer jonne Obolohit alo lagbe
"room-GEN interior-GEN for infra-red light will-be-needed"
'Infra-red light will be needed for the interior of the room'

(28) *ghOrer bhitorkar jonne Obolohit alo lagbe
"room-GEN interior-GEN2 for infra-red light will-be-needed"

In other environments, those that constitute what I will call set B, L forms appear with both *er* and *kar* – in general.

(29) Surjer bhitorer uttap
"sun-GEN interior-GEN heat"
'the heat of the sun's interior'

(30) Surjer bhitorkar uttap
"sun-GEN interior-GEN2 heat"
'the internal heat of the sun'

I say 'in general' because, as the glosses insinuate, there is a slight difference, one which results in occasional semi-acceptabilities.

(31) prithibir Okkho-goti palTanor phOle Sey diner doyrgho chilo Odbhut – kuRi ghOnTa
"earth-GEN axis-motion changing-GEN consequently that day-GEN length was odd – twenty hours"
'Due to change in the earth's axial motion, the length of that day was odd – twenty hours'

(32) ?prithibir Okkho-goti palTanor phOle Sey dinkar doyrgho chilo Odbhut – kuRi ghOnTa
"earth-GEN axis-motion changing-GEN consequently that day-GEN2 length odd was – twenty hours"
'?Due to change in the earth's axial motion, that day the length was odd – twenty hours'

The question mark at (32) is used to indicate pragmatic deviance. While (31) talks about 'the length of that day' being twenty hours, (32) seems to say 'the length was twenty hours that day' – the length of something other than the day, of something not mentioned. I have not yet found stronger clear cases (a case that is stronger but formally less clear is *Sotter* vs. *Sottikarer*, discussed at the very end of this chapter) of the difference between *er* and *kar* showing up in a string that is questionable with one but not with the other.

All K and M forms appear with their appropriate genitive suffixes both in A and in B environments, freely. Only L stems, which have a choice (between *er* and *kar*), show any difference in this regard.

How are the L facts to be handled? Several conceivable methods come to mind. My proposal is as follows. Let us posit the following structures (the category label C stands for Case).

(33) $[_A[_N$ bhitor$][_C \left\{ \begin{matrix} \text{kar} \\ \text{er} \end{matrix} \right\}]]$: in set B environments
(34) $[_N[_N$ bhitor$][_C$ er$]]$: in set A environments

If we take a closer look at the environment sets A and B, we notice that, indeed, set B comprises positions where the form in question may be replaced by a simple adjective without affecting the well-formedness of the construction:

(29) Surjer bhitorkar uttap "sun-GEN interior-GEN2 heat"

(35) Surjer proconDo uttap "sun-GEN intense heat"

– whereas set A comprises positions intolerant of such substitution:

(27) ghOrer bhitorer jonne Obolohit alo lagbe
"room-GEN interior-GEN for infra-red light will-be-needed"

(36) *ghOrer proconDo jonne Obolohit alo lagbe
"room-GEN intense for infra-red light will-be-needed"
(also bad with any AP other than *proconDo* 'intense')

Proposal (33)-(34) easily explains the patterns. What other considerations fortify the proposal?

Suppose we add that (34) is a Phrase Structure while (33) is a morphological pattern; in other words, that there is a PS rule (26-i) that underwrites (34) but no rule A → N C; only a morphological or lexical rule deriving A from N GEN, perhaps. It doesn't quite matter how we obtain (33), provided that it isn't by a Phrase Structure rule.

On the basis of this difference of status between (33) and (34), we can predict that (33) should have more irregularities than (34), since morphology is by and large messier than syntax. Well, (33) does have irregularities of the sort one would expect on this view: *Sotto* 'true, truth', an L stem, has the environment A genitive *Sotter* – absolutely regular – and the environment B genitive *Sotti-kar-er* "true-GEN2-GEN" instead of the expected *Sotto-kar* "true-GEN2". In contrast, there are no known irregularities associated with (34). This is weak confirmation for my ideas about (33) and (34), but it's a start.

Incidentally, I spoke earlier of having found a strong but unclear example of the (31)-(32) phenomenon; what I have just presented, the pair of words derived from *Sotto*, is that strong but formally unclear example. Thus, phrases built around these words have palpably distinct interpretations, e.g. *Sotter jOY* 'truth's victory' and *Sottikarer jOY* 'true victory'. One case where one member of such a pair is a deviant sentence is (37), where the b-example is odd if /Sotto/ is interpreted as 'truth' and completely unacceptable if /Sotto/ is interpreted as 'a truth':

(37) a. eTa EkTa Sottikarer Sal
"this a true-GEN2-GEN shawl"
'this (is) a real shawl'
b. ?eTa EkTa Sotter Sal
"this (is) a truth-GEN shawl"
'?This is a shawl of truth / *This is the shawl of a truth'

Chapter 11

LIST OF RULES

I omit rules of phonology and rules informally stated (with no number assigned) in prose expositions.

From Chapter 2

(1) NP → (NUM′) N
(2) NUM′ → NUML DENOM
(7) NP → N (NUM′); or, (8)
(8) NUM′ Postposing
S.D.: NUM′ – N, where NUM′ contains *Ta* or *khana* or *gulo*
S.C.: 1 2 ➔ 0, 2+1

From Chapter 3

(4) NUML Deletion, obligatory
S.D.: N – NUML – DENOM, with various conditions on the terms
S.C.: 1 2 3 ➔ 1, 0, 3
(32) NP → (DET) (NUM′) N (PLURAL) (CASE)
(36) NP′ → (N PL) NP
(38) Attach *ra* to the word preceding it
(39) Attach *Ta* or *khana* to the word preceding it
(40) Attach Case to the word preceding it
(46) $V \rightarrow \left(\left\{\begin{matrix} N \\ A \end{matrix}\right\}\right) V$
(47) V → (V) V
(60) /ye/-Insertion, obligatory
S.D.: V – V′
S.C.: 1 2 ➔ 1+/ye/, 2

From Chapter 4

(32) Gerund/Participle Rule: $\varnothing_{[[_N V + Wa / no]]_A} (II) = I$

(40) Gerund/Active Participle Rule: $\varnothing_{[[_N V + Wa / no]]_A} (I) = I$

(42) supersedes (32):

Simplified Gerund/Passive Participle Rule: $\varnothing_{N_A} (II) = I$

(50) i. S → NP NP VP V′
ii. VP → NP V
iii. V′ → V T E
iv. Infinitive Reanalysis
S.D.: X – V – V′

S.C: 1 2 3 ➔ 1, 0, 2+3

(63) S.D.: [$_S$ X – V]
S.C.: 1 2 ➔ 1, 2 + [$_T$ ±Future]

(64) S.D.: [$_S$ X – V]
S.C.: 1 2 ➔ 1, 2 + [$_{PERS}$ ±Ego, …]

(65) S.D.: [$_S$ X – [$_V$ Y [$_{PERS}$ –Ego, Z]]]
S.C.: 1 2 ➔ optionally 1, 2+[$_{HON}$ ±Honorific]

(66) S.D.: [$_S$ X – V]
S.C.: 1 2 ➔ 1, 2 + $\left[_M \left\{ \begin{matrix} \text{Present} \\ \text{Past} \\ \text{Imperative} \end{matrix} \right\} \right]$

(67) S.D.: [$_{VP}$ X – V]
S.C.: 1 2 ➔ 1, 2+[$_M$ Present]

(68) S.D.: [$_V$V – V]
S.C.: 1 2 ➔ 1+[$_M$ Past], 2; supersedes (III-60)

(71) S.D.: [$_S$ X – V – Y]
S.C.: 1 2 3 ➔ 1, 2 + NEG; superseded by (72) and (73)

(72) S.D.: [$_{[-S]}$ X – V]
S.C.: 1 2 ➔ 1, NEG + 2

(73) S.D.: [$_S$ X – V]
S.C.: 1 2 ➔ 1, 2 + NEG

From Chapter 5

(82) Negative Perfect Neutralization; superseded by (86)
S.D.: V – PAST – /ch/ – M – T – PERS – NEG
S.C: 1 2 3 4 5 6 7 ➔ 1, 0, 0, 0, 5, 6, 7 + 2

(84) S.D.: [$_V$ (V) – V]
S.C.: 1 2 ➔ 1, 2 + [$_{ASP}$ +Perfect], optional

(85) S.D.: [$_V$ (V) – [$_V$ V ASP]]
S.C.: 1 2 ➔ 1, 2 + V

(86) S.D.: [$_S$ X – +Perf – /ch/ – M – Future – PERS (HON) – NEG]
S.C.: 1 2 3 4 5 6 7
➔ 1, 0, 0, Present, 5, 6, 2

(99) i. S′ → S COMP
ii. COMP → ±INT

From Chapter 6

(34) Complementizer Preposing
S.D.: X – Y – Z – COMP
S.C.: 1 2 3 4 ➔ 1, 2+4, 3, 0

(50) X^m → nX^m

(56) S → (S′) (PP) NP (PP) (NP) V

(75) Conjunctionization (obligatory)
S.D.: X – CONJ – S – +CONJ
S.C.: 1 2 3 4 ➔ 1, 4, 3, 0

(84) EMP Preposing
S.D.: X – X^o – Y – EMP
S.C.: 1 2 3 4 ➔ 1, 2 + 4, 3, 0

(88) Sentential EMP Placement (root transformation)
S.D.: X – [–S] – Y
S.C.: 1 2 3 ➔ 1, 2 + EMP, 3

(89) Nominal EMP Placement (NP-cyclic transformation)
S.D.: X – NP – X
S.C.: 1 2 3 ➔ 1, 2 + EMP, 3

From Chapter 7

(16) Phoric Determiner Postposing
S.D.: +DET +PHOR – X – Y
S.C.: 1 2 3 ➔ 0, 2, 1 + 3

From Chapter 8

(67) Interrogative Construal
S.D.: an S′ which dominates, without any lower S′ also dominating, zero or more +INT determiners and a +INT complementizer
S.C.: construe all the determiners with the complementizer

(68) Open Specifier Construal
S.D.: an S′ which dominates, without any lower S′ also dominating, zero or more α INT determiners and an α INT complementizer
S.C.: construe all the determiners with the complementizer

(71) Relative Binding
S.D.: an S′, whose complementizer is -INT and with which a construal ((y_1, y_2, …, y_n), x) is associated, embedded in a larger S′
S.C.: make it so that each variable is bound by at least one antecedent to the right of the embedded S′ and any single antecedent in the matrix directly binds at most one variable

(77) Interrogative Binding
S.D.: an S′ whose complementizer is +INT and with which a construal ((y_1, y_2, …, y_n), x) is associated
S.C.: enlarge the construal to include an e binding x

From Chapter 9

(25) Alternative Conjunctionization (obligatory)
S.D.: [$_\alpha$ X – +COMP +CONJ +OPEN +VBL +INT – Y – +CONJ – Z]
S.C.: 1 2 3 4 5 ➔ 1, 0, 3, 2, 5
Condition: α is an indirect question

BIBLIOGRAPHY

Altham, J. E. J., & Tennant, N. W. (1975). Sortal quantification. In E. L. Keenan (Ed.), *Formal Semantics of Natural Language* (pp. 46-58). Cambridge: Cambridge University Press.

Anderson, S. R. (1977). Comments on the paper by Wasow. In A. Akmajian, P. Culicover & T. Wasow (Eds.), *Formal Syntax* (pp. 361-377). New York: Academic Press.

Aronoff, M. (1976). *Word Formation in Generative Grammar*. Cambridge, Massachusetts: MIT Press.

Bachenko, J. (1976). *Some Semantic Rules in the Formal Description of English Pseudo-Cleft Sentences*. Ph.D. dissertation. New York University, New York.

Bachenko, J., & Teller, V. (forthcoming). Ambiguity, dependency, and inviolability in pseudo -cleft sentences.

Baker, C. L. (1970). Notes on the description of English questions: the role of an abstract question morpheme. *Foundations of Language, 6*, 197-219.

Baker, C. L. (1978). *Introduction to Generative-Transformational Syntax*. Englewood Cliffs, N.J.: Prentice Hall.

Baltin, M. R. (1978). *Toward a Theory of Movement Rules*. Ph.D. dissertation. Massachusetts Institute of Technology, Cambridge.

Basu, R. (1973). *Calantikā*. Calcutta: M.C. Sarkar.

Belnap, N. D. Jr., & Steel, T. B. Jr. (1976). *The Logic of Questions and Answers*. New Haven, Conn: Yale University Press.

Boër, S. E. (1978). 'Who' and 'whether': towards a theory of indirect question clauses. *Linguistics and Philosophy, 2*, 307-345.

Bose, K. D. (n.d.) *The Essentials of Bengali Grammar*. Calcutta: Mitra & Ghosh.

Brame, M. (1976). *Conjectures and Refutations in Syntax and Semantics*. New York: Elsevier North-Holland.

Brame, M. (1977). Alternatives to the tensed-S and specified subject conditions. *Linguistics and Philosophy, 1*, 381-411.

Brame, M. (1979). *Base Generated Syntax*. Seattle: Noit Amrofer.

Bresnan, J. (1970). On complementizers: toward a syntactic theory of complement types. *Foundations of Language, 6*, 297-321.

Bresnan, J. (1973). Sentence stress and syntactic transformations. In K. J. J. Hintikka, J. M. E. Moravcsik, & P. Suppes (Eds.), *Approaches to Natural Language* (pp. 3-47). Dordrecht: Reidel.

Bresnan, J. (1976a). On the form and functioning of transformations. *Linguistic Inquiry*, *7*, 3-40.

Bresnan, J. (1976b). Evidence for a theory of unbounded transformations. *Linguistic Analysis*, *4*, 353-394.

Bresnan, J. (1978). A realistic theory of transformational grammar. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic Theory and Psychological Reality* (pp. 1-59). Cambridge: MIT Press.

Chatterji, S. K. (1926). *The Origin and Development of the Bengali Language*. 1971 reissue, 3 vols. London: George Allen & Unwin.

Chatterji, S. K. (1968). *Saral Bhāṣāprakāś Bāṅlā Byākaraṇ*. Calcutta: Vaksahitya.

Chattopadhyay, J. (1976a). *Complement Constructions in Bengali*. M.Litt. dissertation. University of Delhi.

Chattopadhyay, J. (1976b). Complementizer deletion in Bengali as a global rule. *Papers in Linguistic Analysis*, *1*, 1-10.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.

Chomsky, N. (1970). Remarks on nominalization. In R. A. Jacobs, & P. S. Rosenbaum (Eds.), *Readings in English Transformational Grammar* (pp. 184-221). Waltham, Mass.: Ginn and Company.

Chomsky, N. (1972). *Language and Mind*. New York: Harcourt, Brace & Jovanovich.

Chomsky, N. (1977a). *Essays on Form and Interpretation*. New York: North Holland.

Chomsky, N. (1977b). On *wh*-movement. In A. Akmajian, P. Culicover, & T. Wasow (Eds.), *Formal Syntax* (pp. 71-132). New York: Academic Press.

Chomsky, N. (1978). On binding. Manuscript. [Shortly after I submitted this thesis, the paper, which I had consulted in its manuscript form, appeared under the same title in *Linguistic Inquiry*, *11:1*, 1-46, in 1980.]

Chomsky, N., & Halle, M. (1968). The Sound Pattern of English. New York: Harper & Row.

Chomsky, N., & Lasnik, H. (1977). Filters and control. *Linguistic Inquiry, 8,* 425-504.

Cohen, F.S. (1929). What is a question? *The Monist, 39,* 355-364.

Das, J. M. (1916). *Bāṅgālā Bhāṣār Abhidhān*. Calcutta: The India Publishing House.

Dasgupta, P. (1977). The internal grammar of the compound verb in Bangla. *Indian Linguistics, 38,* 68-85.

Dasgupta, P. (1979). The mantissa of indefinite words in Bangla. *International Journal of Dravidian Linguistics, 8,* 46-51.

Dasgupta, P. (forthcoming in 1981). The Bangla verb stem formative -*a*. In L. Levine, & F. Ulfers (Eds.), *Festschrift for Robert Fowkes*. [Editorial note dated 2020: After the plans for the Robert Fowkes festschrift fell through, this paper, originally written in 1977, was published as: Dasgupta, P. (2000). The verb stem formative /a/ in Bangla. In R. E. Asher, & R. Harris (Eds.) *Linguisticoliterary: A festschrift for Professor D.S. Dwivedi* (pp. 115-125). Delhi: Pilgrim. A conceptually updated analysis of the phenomena was provided in: Dasgupta, P. (2001). On a vowel template asymmetry in Bangla verbs. In A. Abbi, R. S. Gupta, & A. Kidwai (Eds.), *Linguistic Structure and Language Dynamics in South Asia: Papers from Proceedings of SALA XVIII Roundtable* (pp. 164-181). Delhi: Motilal Banarsidass.]

Dasgupta, P. (1982). The phonology of the Bangla verb. *Indian Linguistics, 43,* 17-28. [A revised version of the 1972 manuscript 'The phonology of the Western Bangla verb'.]

Delorme, B., & Dougherty, R. C. (1972). Appositive NP constructions: *we, the men; we men; I, a man;* etc. *Foundations of Language, 8,* 2-29.

Dixon, R. M. W. (1966). Relative clauses and possessive phrases in two Australian languages. *Language, 45,* 35-44.

Dougherty, R. C. (1969). An interpretive theory of pronominal reference. *Foundations of Language, 5,* 488-519.

Dougherty, R. C. (1970). A grammar of coördinate conjoined structures: I. *Language, 46,* 850-898.

Dougherty, R. C. (1971). A grammar of coördinate conjoined structures: II. *Language, 47,* 298-339.

Emonds, J. (1970). *Root and Structure-Preserving Transformations*. Ph.D. dissertation. Massachusetts Institute of Technology, Cambridge.

Emonds, J. (1973). Evidence that indirect object movement is a structure-preserving rule. In M.

Gross, M. Halle, & M.-P. Schützenberger (Eds.), *The Formal Analysis of Natural Languages* (pp. 73-87). The Hague: Mouton.

Emonds, J. (1976). *A Transformational Approach to English Syntax*. New York: Academic Press.

Epée, R. (1976). A counterexample to the Q replacement and COMP substitution universals. *Linguistic Inquiry, 7,* 677-686.

Ferguson, C. F. (1945). *The Phonology and Morphology of Standard Colloquial Bengali*. Ph.D. dissertation. University of Pennsylvania.

Fiengo, R. (1974). *Semantic Conditions on Surface Structure*. Ph.D. dissertation. Massachusetts Institute of Technology, Cambridge.

Fiengo, R. (1978). *Surface Structure*. Manuscript.

Fillmore, C. (1977). The case for case reopened. In P. Cole, & J. M. Sadock (Eds.), *Syntax and Semantics, Vol. 8, Grammatical Relations* (pp. 59-61). New York: Academic Press.

Forbes, D. 1861. *A Grammar of the Bengali Language*. London: Sampson, Low, Marston & Co., Publishers to the India Office.

Grinder, J. (1971). Double indices. *Linguistic Inquiry*, *2*, 572.

Halitsky, D. (1974). Deep structure appositive and complement NPs. *Language*, *50*, 446-454.

Halitsky, D. (1975). Left branch S's and NP's in English: a bar notation analysis. *Linguistic Analysis, 1*, 279-296.

Helke, M. (1975). The psychological reality of lexical items. Manuscript. [Consulted as a manuscript while drafting this thesis in 1979 and finalizing it in 1980. The paper was published with minor revisions in 1979, but I was unaware of this while writing: Helke, M. (1979). The psychological reality of lexical items. In G. Bedell, E. Kobayashi, & M. Muraki (Eds.), *Explorations in Linguistics: Papers in Honour of Kazuko Inoue* (pp. 151-200). Tokyo: Kenkyusha.]

Hinds, J. (1974). On the status of the VP node in Japanese. Manuscript.

Hornstein, N. (1975). S and the X´ convention. *Montreal Working Papers in Linguistics*, *4*, 35-71.

Jackendoff, R. (1977). *X´ Syntax: a Theory of Phrase Structure*. Cambridge, Mass.: MIT Press.

Jakobson, R., Halle, M., & Fant, G. (1952). *Preliminaries to Speech Analysis*. Cambridge, Mass.: MIT Press.

Jespersen, O. (1965). *The Philosophy of Grammar*. New York: Norton.

Kar, B. (1376 B.S.) [= 1969-70 C.E.] *Saṅginī*. Calcutta: Mitra and Ghosh.

Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy, 1*, 3-44.

Katz, J. J., & Postal, P. M. (1964). *An Integrated Theory of Linguistic Descriptions*. Cambridge, Mass.: MIT Press.

Kaufman, E. S. (1974). Navajo spatial enclitics: a case for unbounded rightward movement. *Linguistic Inquiry, 5*, 507-533.

Klaiman, M. H. (1977). Bengali syntax: possible Dravidian influences. *International Journal of Dravidian Linguistics, 6*, 303-317.

Kohrt, M. (1975). A note on bounding. *Linguistic Inquiry, 6*, 167-171.

Kuno, S., & Robinson, J. (1972). Multiple WH questions. *Linguistic Inquiry, 3*, 463-488.

Kuroda, S.-Y. (1968). English relativization and certain related problems. *Language, 44*, 244-266.

Kuroda, S.-Y. (1974, 1976, 1977). Pivot-independent relativization in Japanese: I. *Papers in Japanese Linguistics, 3*.59-93; II, *PJL, 4*.85-96; III, *PJL, 5*.157-179.

Lees, R. (1960). *The Grammar of English Nominalizations*. The Hague: Mouton.

Lyons, J. (1969). *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.

McCawley, J. (1977). Lexicographic notes on English quantifiers. *CLS, 13*, 372-384.

Malone, J. (1977). Auto-referential determination. *CLS, 13*, 352-356.

Matthews, P. (1974). *Morphology*. Cambridge: Cambridge University Press.

May, R. (1977). *The Grammar of Quantification*. Ph.D. dissertation. Massachusetts Institute of Technology, Cambridge.

Misra, V. N. (1966). *Studies in the Descriptive Method of Panini*. The Hague: Mouton.

Mostowski, A. (1957). On a generalization of quantifiers. *Fundamenta Mathematicae, 44*, 12 -36.

Ohira, E. L. (1975). The surface branching constraint. *Presented at 20th Conference of*

International Linguistic Association, New York.

Partee, B. H. (1973). The syntax and semantics of quotation. In S. R. Anderson, & P. Kiparsky (Eds.), *A Festschrift for Morris Halle* (pp. 410-418). New York: Holt, Rinehart & Winston.

Partee, B. H. (1975). Montague grammar and transformational grammar. *LI*, *6*, 203-300.

Radford, A. (1975). Pseudo-relatives and the unity of subject raising in French. *Archivum Linguisticum* (new series) *6*, 32-64.

Ray, P. S., Hai, M. A., & Ray, L. (1966). *Bengali Language Handbook*. Washington, D.C.: Center of Applied Linguistics.

Reed, A. (1975). *The Structure of Relative Clauses*. Ph.D. dissertation. Brandeis University.

Reichenbach, H. (1947). *Elements of Symbolic Logic*. New York: The Free Press.

Ross, J. R. (1967). Constraints on Variables in Syntax. Ph.D. dissertation. Massachusetts Institute of Technology, Cambridge.

Rothenberg, M. (1972). Les propositions relatives adjointes en français. *Bulletin de la Société de Linguistique de Paris*, *67*, 175-213.

Sag, I. (1976). *Deletion and Logical Form*. Ph.D. dissertation. Massachusetts Institute of Technology, Cambridge.

Schachter, P. (1977). Constraints on coördination. *Language*, *53*, 86-103.

Schwartz, A. (1972). The VP-constituent of SVO languages. In J. P. Kimball (Ed.), *Syntax & Semantics: Vol. I* (pp. 213-234). New York: Academic Press.

Seegmiller, S. M. (1974). *Lexical Insertion in a Transformational Grammar*. Ph.D. dissertation. New York University, New York.

Selkirk, E. (1972). *The Phrase Phonology of English and French*. Ph.D. dissertation. Massachusetts Institute of Technology, Cambridge.

Singh, U. N. (1976). Negation in Bengali and the order of constituents. *Indian Linguistics*, 37, 295-303.

Tagore, R. (1938). *Bāṅlā Bhāṣā-paricay*. Calcutta: Visvabharati.

Wasow, T. (1972). *Anaphoric Relations in English.* Ph.D. dissertation. Massachusetts Institute of Technology, Cambridge.

Wasow, T. (1977). Transformations and the lexicon. In A. Akmajian, P. Culicover, & T. Wasow (Eds.), *Formal Syntax* (pp. 327-360). New York: Academic Press.

Whitehead, A. N. (1933). *Adventures of Ideas*. 1948 reissue. Harmondsworth, England: Penguin Books.

Woisetschläger, E. (1976). On how useful a construct the left-branch condition is. In A. Ford, J. Reighard, & R. Singh (Eds.), *Proceedings of the Sixth Conference of the North-Eastern Linguistics Society. Montréal Working Papers in Linguistics*.

Zbavitel, D. (1970). *Non-finite Verbal Forms in Bengali*. Prague: The Oriental Institute in the Publishing House of the Czechoslovak Academy of Sciences.